Parameter Estimation in EnKF: Surface Fluxes of Carbon, Heat, Moisture and Momentum

Ji-Sun Kang, **Eugenia Kalnay**, Takemasa Miyoshi, Junjie Liu, Inez Fung, Kayo Ide and many collaborators at the University of Maryland **Weather-Chaos** group

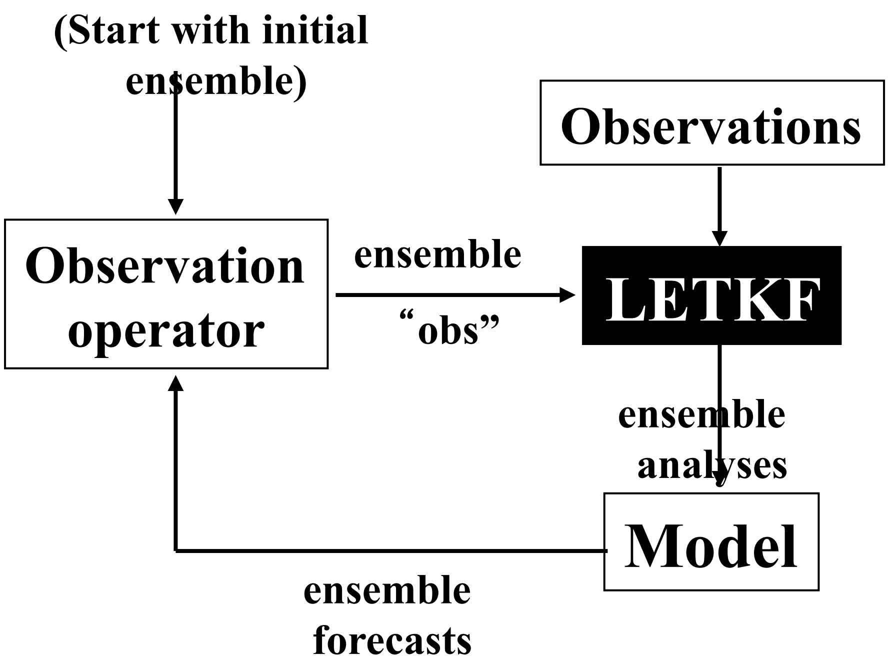
Outline

- A few **recent advances** in LETKF
  - Running in Place (Yang et al, Penny et al)
  - Effective assimilation of precipitation (Lien et al)
  - Ensemble Forecast Sensitivity to Observations (EFSO)

  (Kalnay et al, Ota et al, Daisuke Hotta, thanks to JCSDA!)
- **Parameter estimation** with LETKF allows us to estimate surface fluxes.
- **Simultaneous assimilation** of carbon and meteorological observations
- **Advanced methods:** “variable localization”, vertical localization based on processes, additive and multiplicative adaptive localization)
- **Are short or long assimilation windows better?** We use 6hr windows

Results:
- **Carbon cycle** data assimilation OSSE **successful** results with **LETKF-C**
- Estimation of **surface heat and moisture fluxes**
- Estimation of **wind stress** in addition to SHF and LHF
- Plans

4D-Local Ensemble Transform Kalman Filter
(Ott et al, 2004, Hunt et al, 2004, 2007)

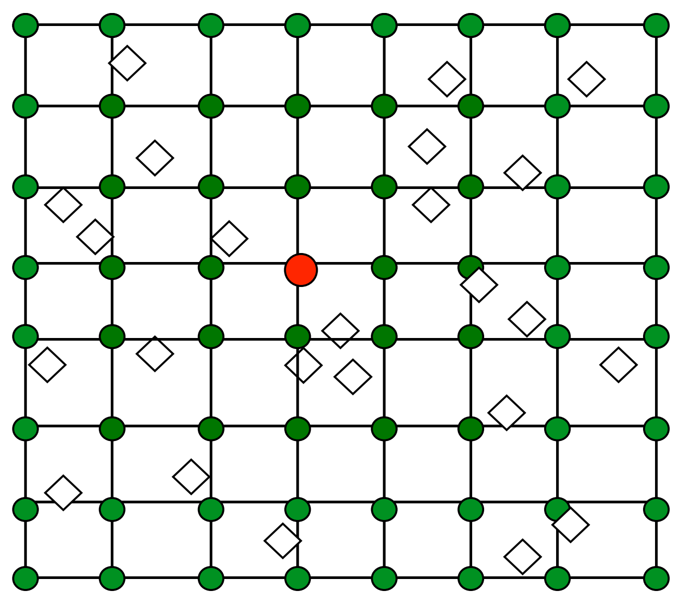


- **Model independent (black box)**
- **No adjoint needed**
- **4D LETKF extension**
- **Obs. assimilated simultaneously at each grid point**
- **LETKF computes the weights for the ensemble forecasts explicitly**

Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

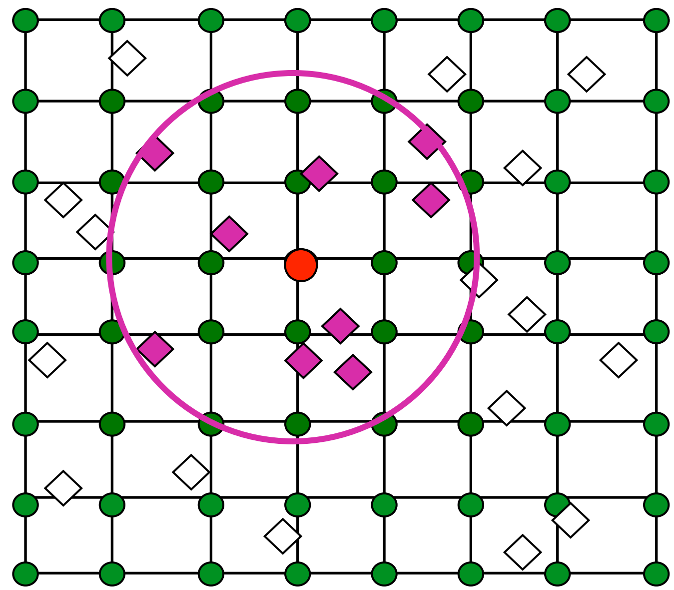


Localization based on observations

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

All observations (purple diamonds) within the local region are assimilated



The LETKF algorithm can be described in a single slide!

Local Ensemble Transform Kalman Filter (LETKF)

Globally:

Forecast step: $\mathbf{x}^b_{n,k} = M_n\left(\mathbf{x}^a_{n-1,k}\right)$

Analysis step: construct $\mathbf{X}^b = \left[\mathbf{x}^b_1 - \overline{\mathbf{x}}^b \mid \ldots \mid \mathbf{x}^b_K - \overline{\mathbf{x}}^b\right];$

$$\mathbf{y}^b_i = H(\mathbf{x}^b_i);\ \mathbf{Y}^b_n = \left[\mathbf{y}^b_1 - \overline{\mathbf{y}}^b \mid \ldots \mid \mathbf{y}^b_K - \overline{\mathbf{y}}^b\right]$$

Locally: Choose for each grid point the observations to be used, and compute the local analysis error covariance and perturbations in ensemble space:

$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^T\mathbf{R}^{-1}\mathbf{Y}\right]^{-1};\ \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

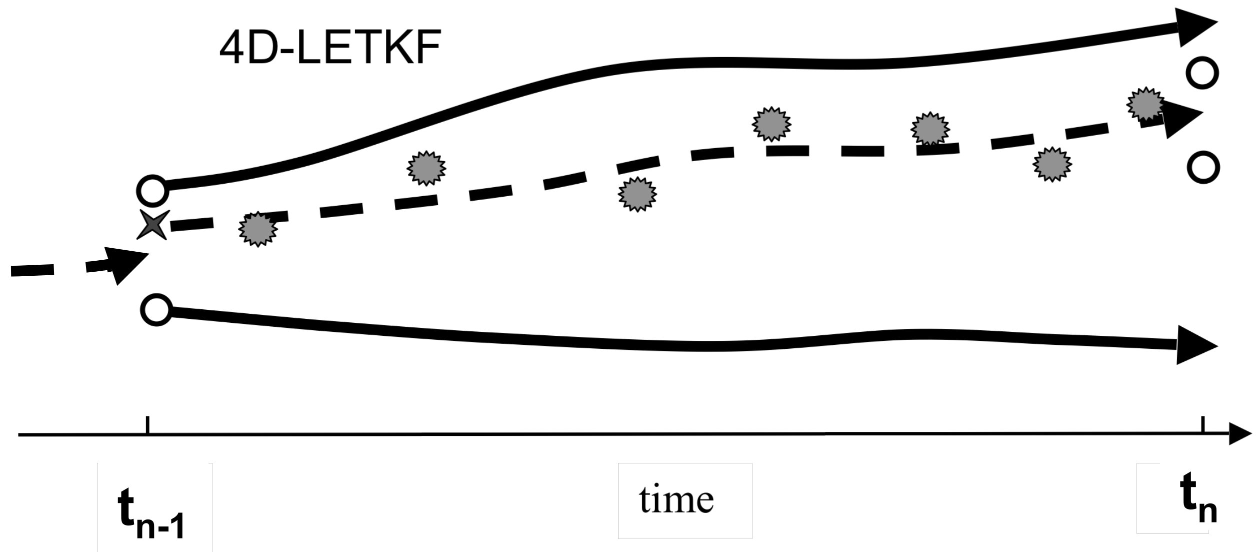
Analysis mean in ensemble space: $\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$

and add to $\mathbf{W}^a$ **to get the analysis ensemble in ensemble space.**

The new ensemble analyses in model space are the columns of $\mathbf{X}^a_n = \mathbf{X}^b_n\mathbf{W}^a + \overline{\mathbf{x}}^b$ **. Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights** $\overline{\mathbf{w}}^a$ **and perturbation analysis weight matrices** $\mathbf{W}^a$

These weights multiply the ensemble forecasts.

No-cost LETKF smoother (×): apply at t_{n-1} the same weights found optimal at t_n. It works for 3D- or 4D-LETKF



The no-cost smoother makes possible:

- ✓ Quasi Outer Loop (QOL)
- ✓ “Running in place” (RIP) for faster spin-up
- ✓ Use of future data in reanalysis
- ✓ Ability to use longer windows and nonlinear perturbations

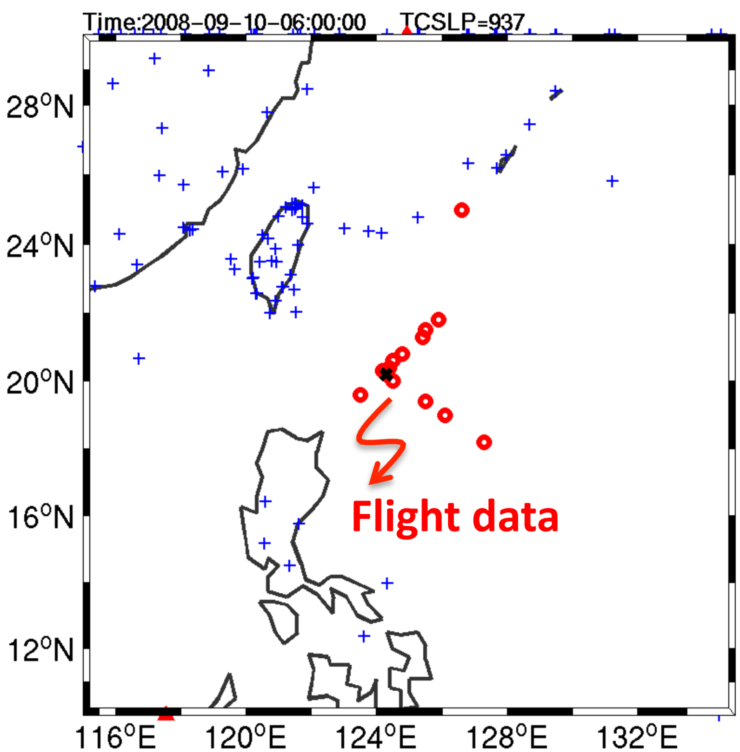
Kalnay & Yang, 2010, Yang et al, 2012

Promising new tools for the LETKF (1)

1. Running in Place (Kalnay and Yang, QJ 2010, Yang, Kalnay, and Hunt, MWR, 2012)

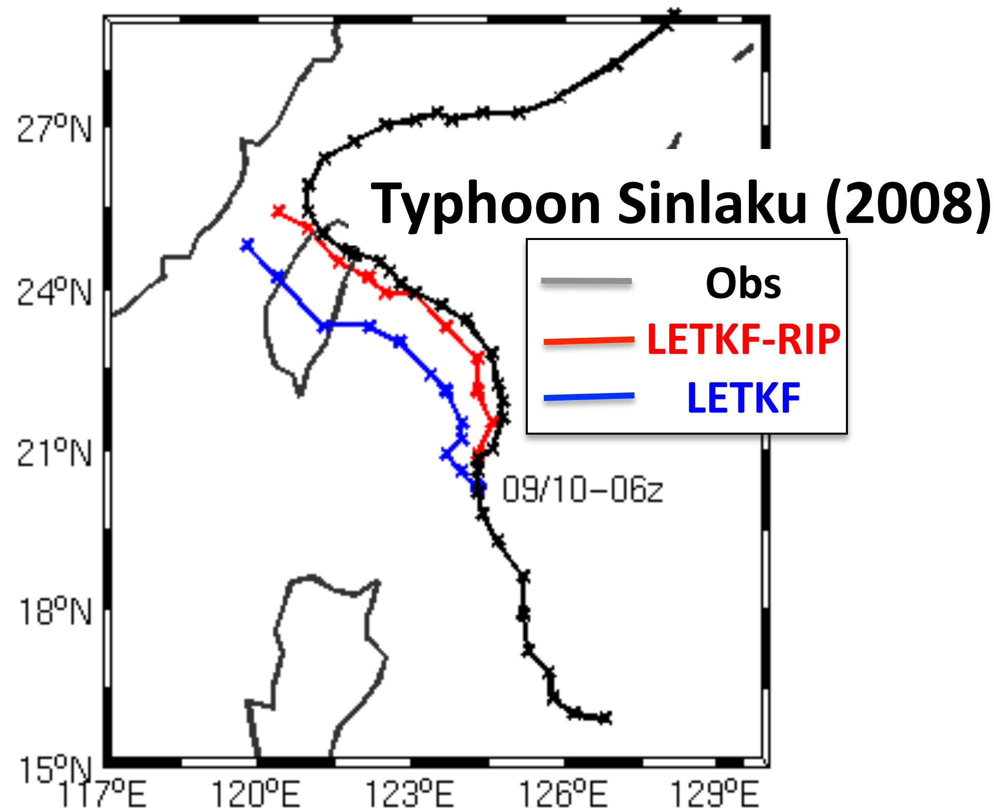
- It extracts more information from observations by using them more than once (sometimes considered a mortal sin!).
- Useful during spin-up (e.g., hurricanes and tornados).
- It uses the "no-cost smoother", Kalnay et al., Tellus, 2007b.
- Typhoon Sinlaku (Yang et al., 2012)
- 7-years of Ocean Reanalysis (Penny, 2011, Penny et al., 2013)

LETKF-RIP with real observations (Typhoon Sinlaku, 2008)



SYNOP(+),SOUND(△),
DROPSONDE(○),
Typhoon center (X)

3-day forecast

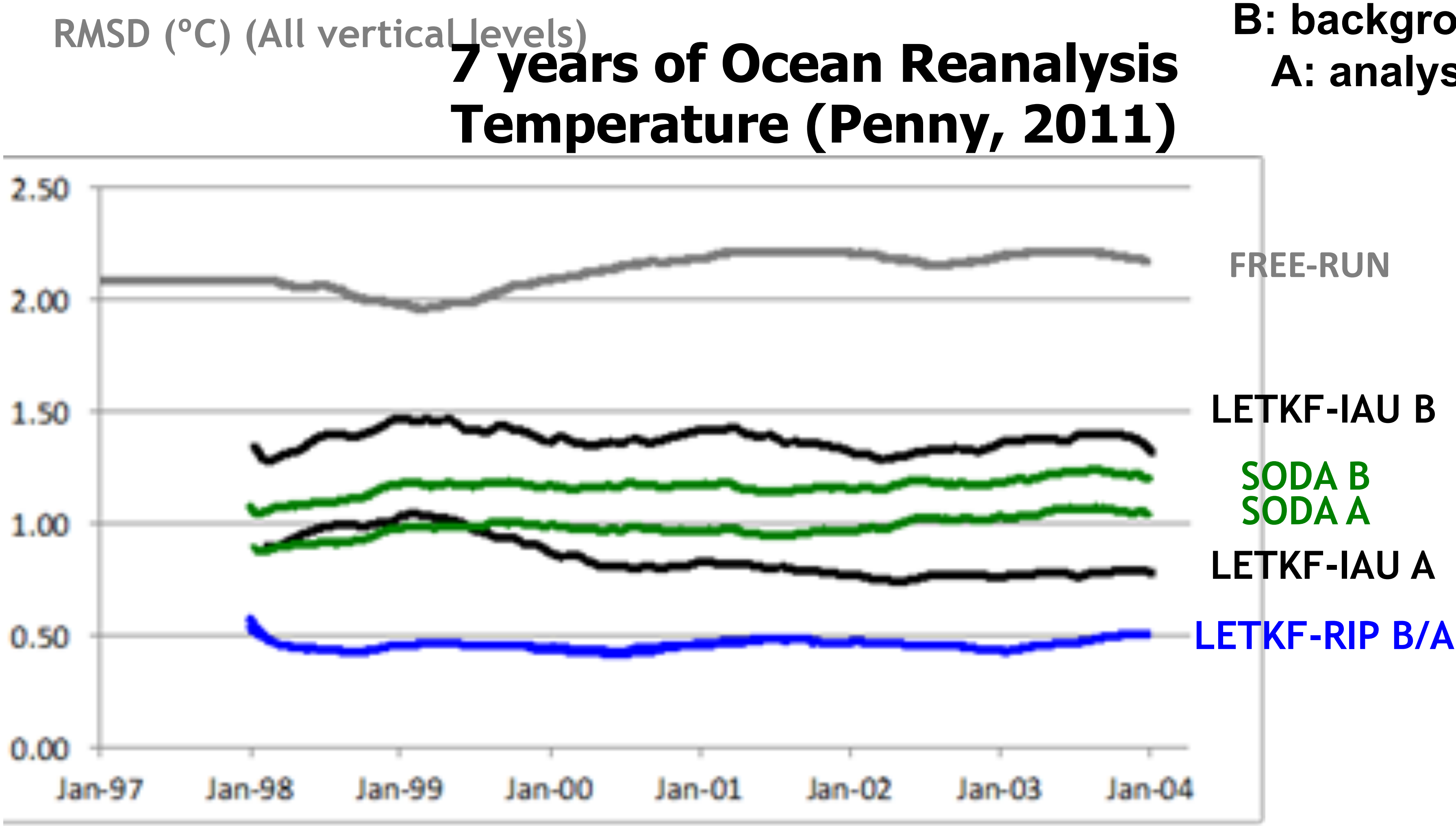


RIP uses better the "limited observations"!

Courtesy of Prof. Shu-Chih Yang (NCU, Taiwan)

7 years of Ocean Reanalysis Temperature (Penny, 2011)

B: background
A: analysis

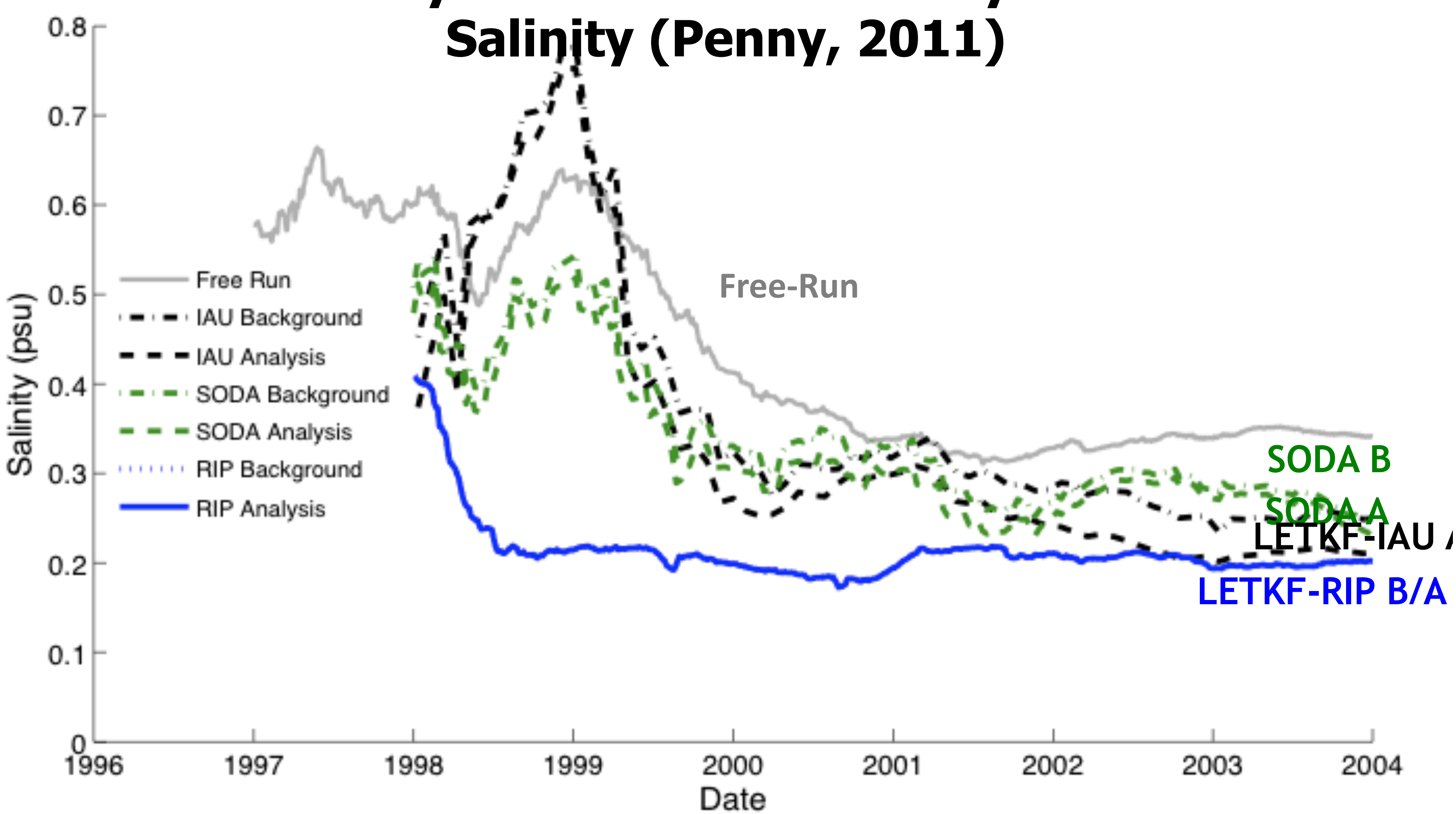


Global RMS(O-F) of Temperature (°C), 12-month moving average
LETKF with IAU, SODA and LETKF with RIP

RMSD (psu) (All vertical levels)

B: background
A: analysis

7 years of Ocean Reanalysis Salinity (Penny, 2011)



Global RMS(O-F) of Salinity (psu), 12-month moving average
LETKF with IAU, SODA and LETKF with RIP

Promising new tools for the LETKF (2)

2. Effective assimilation of Precipitation (**Guo-Yuan Lien**, E. Kalnay and T. Miyoshi, 2013)

- Assimilation of precipitation has generally failed to improve forecasts beyond a day.
- A new approach deals with non-Gaussianity, and assimilation of both zero and non-zero precipitation.
- Rather than changing moisture to force the model to rain as observed, the LETKF changes the potential vorticity.
- But LETKF needs Gaussian errors.
- So, we tried converting precip into a Gaussian distribution.
- The model now "remembers" the assimilation, so that that medium range forecasts are improved in the OSSEs.

How do we transform precipitation *y* to a Gaussian y_{transf}? (Lien et al. 2013)

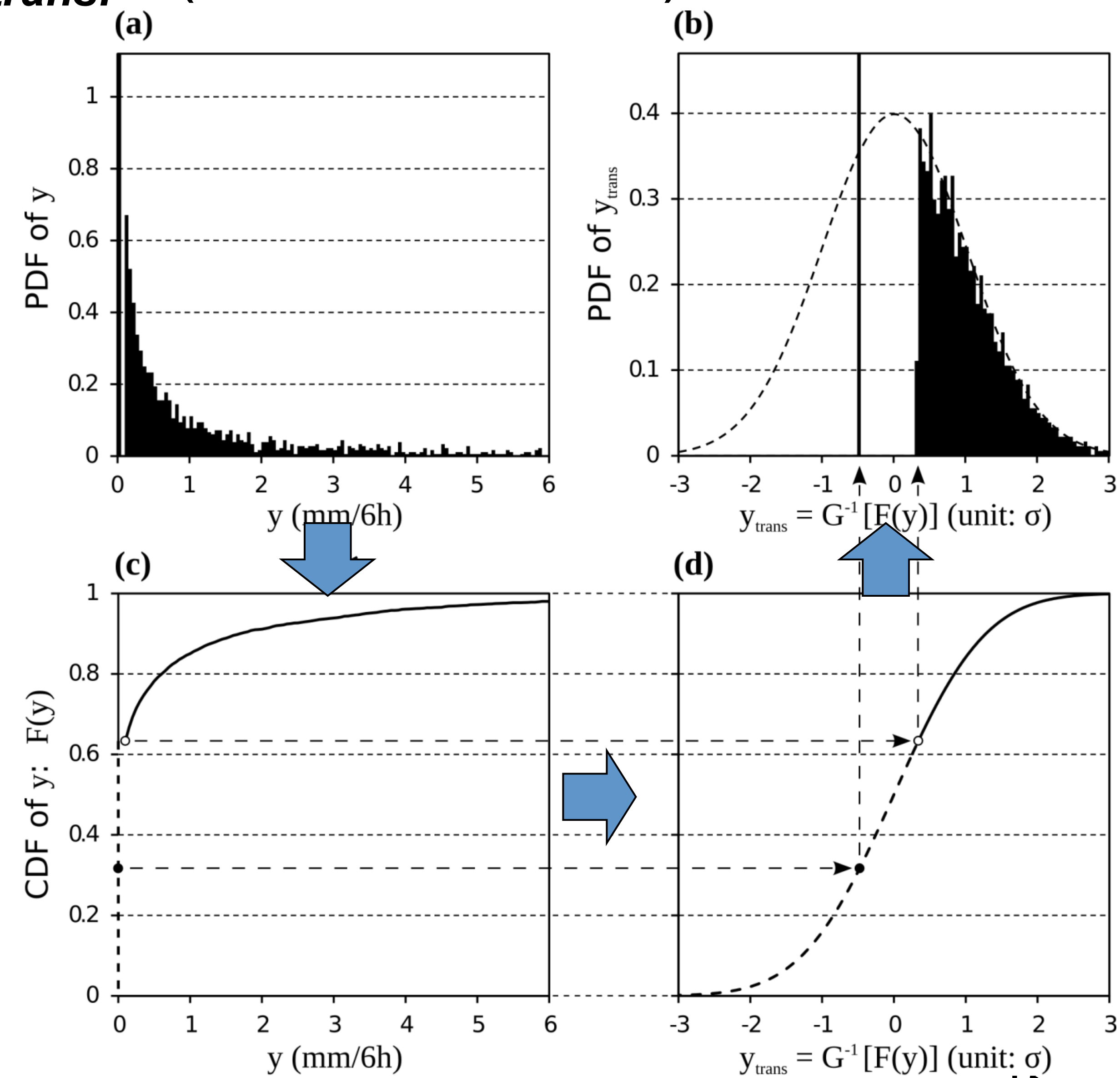
Start with pdf of *y*=rain at every grid point.

"No rain" is like a delta function that we cannot transform.

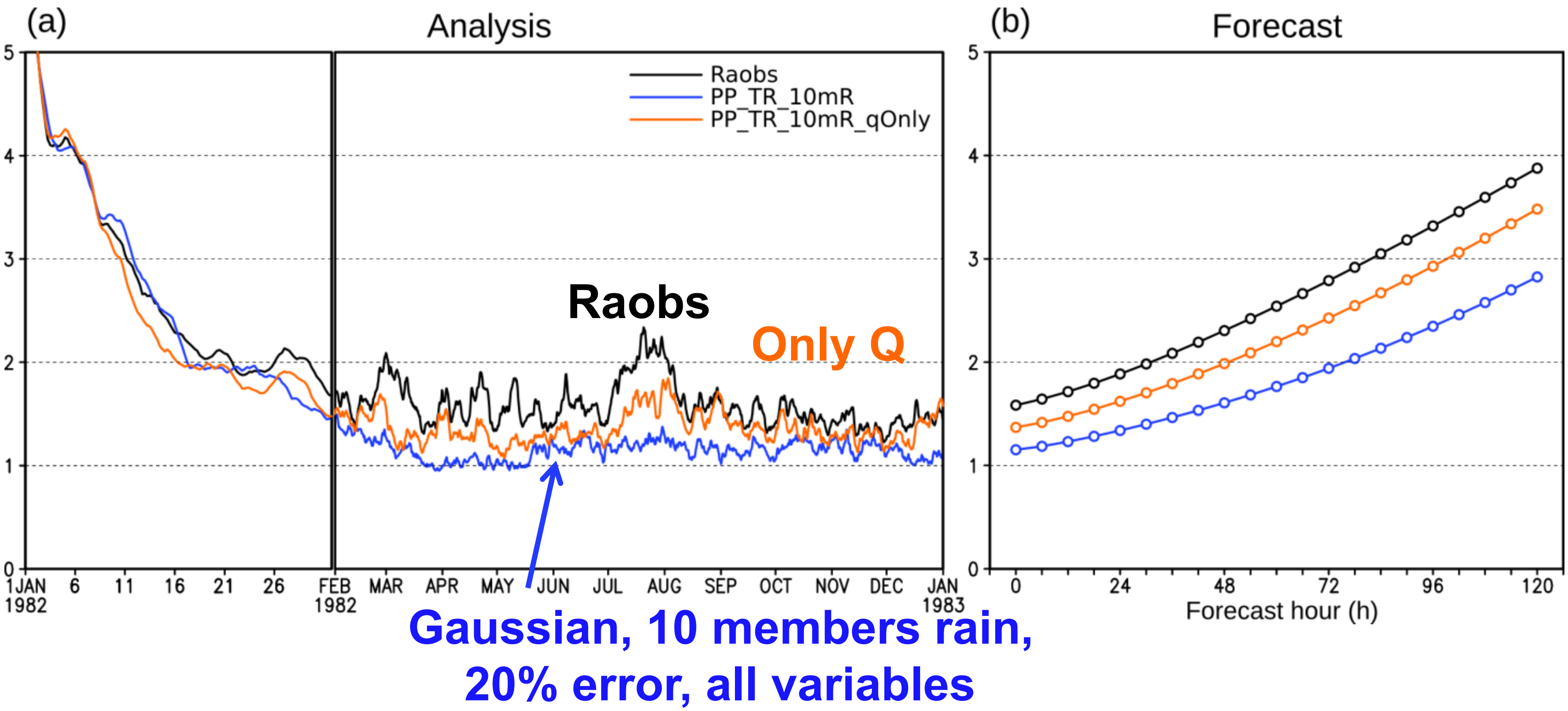
We assign all "no rain" to the median of the no rain CDF.

We found this works as well as more complicated procedures.

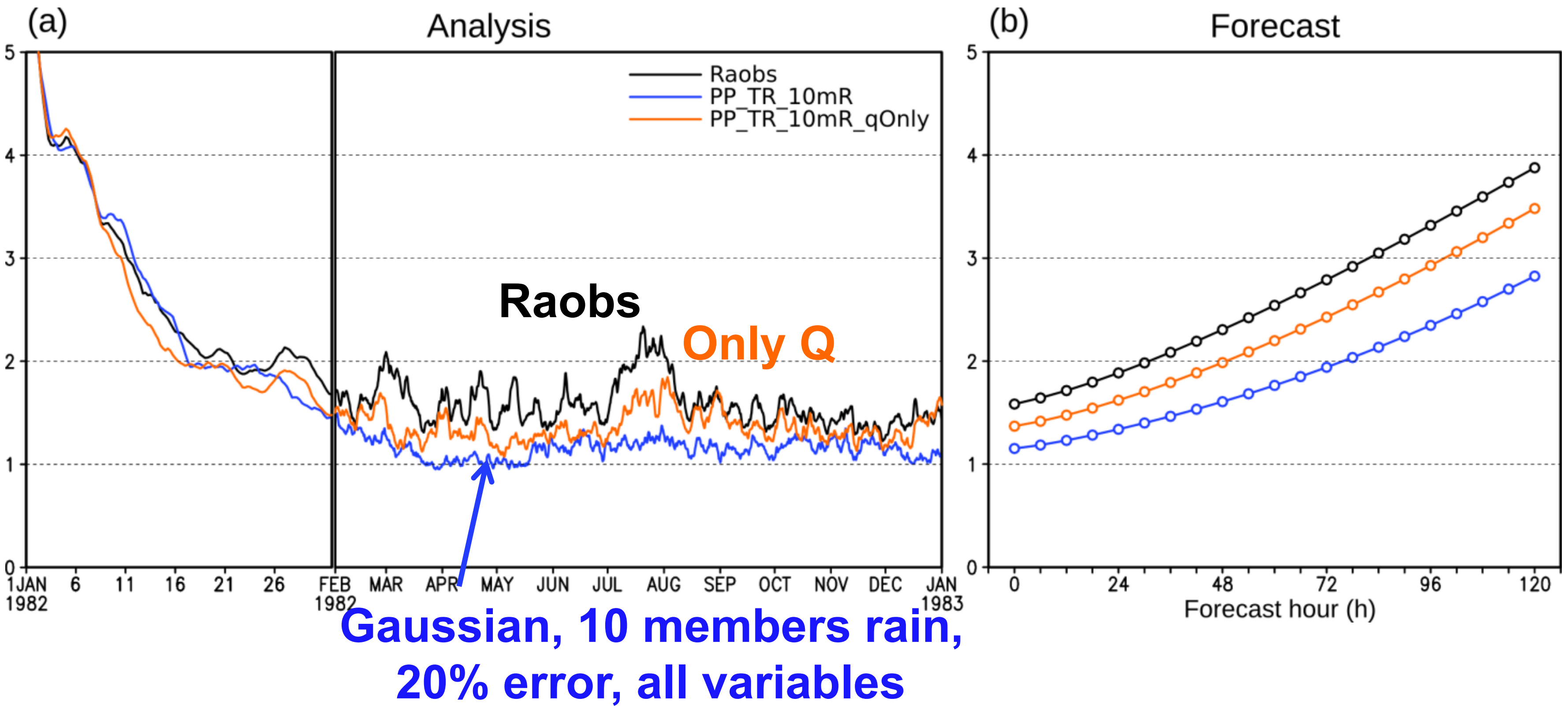
It allows to assimilate both rain and no rain.



$$G^{-1}(x) = \sqrt{2}\,erf^{-1}(2x-1)$$



- **Main result:** **with at least 10 ensemble members raining in order to assimilate an obs, updating all variables (including vorticity), with Gaussian transform, and rather accurate observations (20% errors), the analyses and forecasts are much improved!**
- **Updating only Q is much less effective.**
- **The 5-day forecasts maintain the advantage!**



- Plans (Lien et al., 2014):

- OSSEs with imperfect model: GFS nature, SPEEDY model.

- Assimilate into GFS the TMPA (TRMM+) global precipitation, in preparation for the new PMM system.

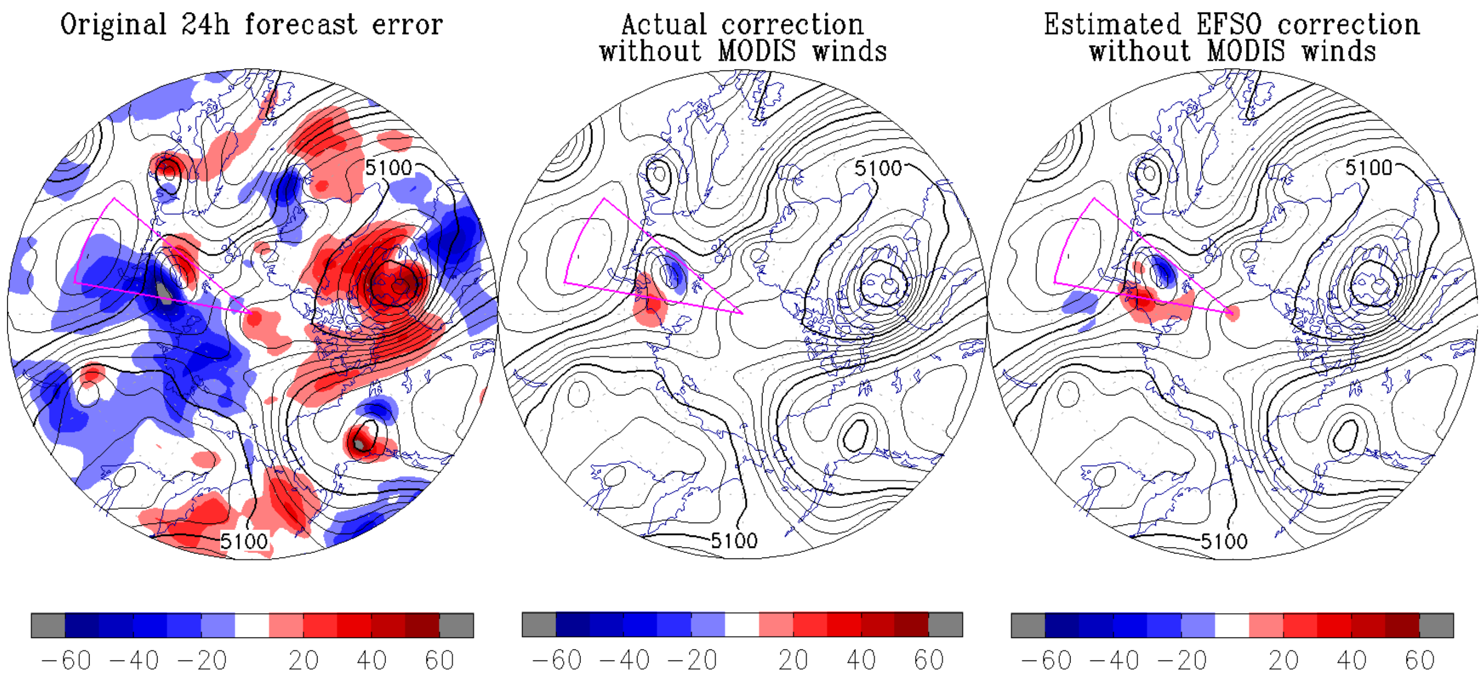
Promising new tools for the LETKF (3)

3. Forecast Sensitivity to Observations and proactive QC

(with Y Ota, T Miyoshi, J Liu, J Derber, D Hotta)

- A simpler, more accurate formulation for the Ensemble Forecast Sensitivity to Observations (EFSO, Kalnay et al., 2012, Tellus).
- Ota et al., 2013 (Tellus) tested it with the NCEP EnSRF-GFS operational system using all operational observations.
- The results obtained comparing the impact of **all** obs. are similar to Langland and Baker (2004) and Gelaro and Zhu.
- Allows to identify "bad observations" after 12 or 24hr, and then repeat the data assimilation without them: "**proactive QC**".

“Proactive QC”: Bad observations can be identified by EFSO and withdrawn from the data assimilation



After identifying MODIS polar winds producing bad 24 hr regional forecasts, the withdrawal of these winds reduced the forecast errors by 39%, as projected by EFSO.

Promising new tools for the LETKF (4)

4. Estimation of surface fluxes as evolving parameters

(Kang et al., 2011, JGR, Kang et al., 2012, JGR)

- important for the carbon cycle
- surface fluxes of heat, moisture, and momentum
- eventually for coupled data assimilation

This is the rest of the talk: Ji-Sun Kang*

with E. Kalnay, J. Liu and Inez Fung

*** Now at KIAPS**

(Korean Institute for Atmospheric Prediction Systems)

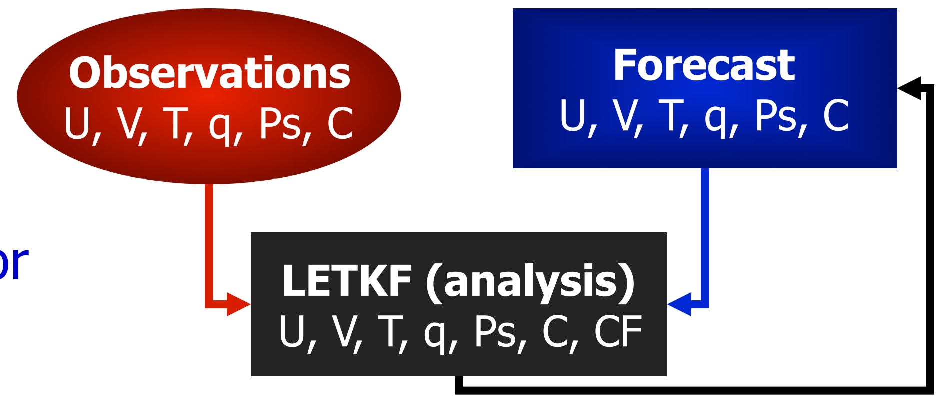
UMD-UCB LETKF-C System

Parameter estimation:
state vector augmentation

$$\mathbf{X}^b = \begin{bmatrix} \mathbf{X} \\ \mathbf{CF} \end{bmatrix}$$

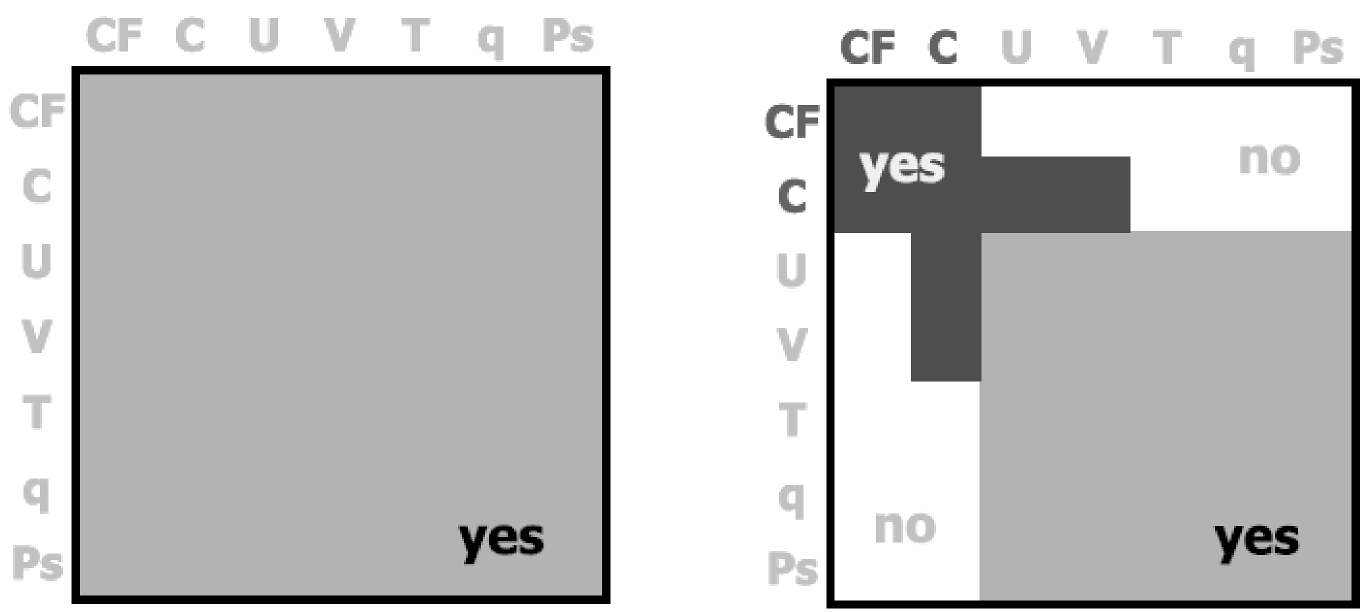
: model state vector (U, V, T, q, Ps, C)

: surface CO_2 flux



  - Append **CF** (surface CO_2 fluxes)
  - Update **CF** as part of the data assimilation process
- **Simultaneous** assimilation of carbon and meteorological variables
  - **Multivariate** analysis with **a localization of the variables** (Kang et al., 2011)
  - Update all variables (including **CF**) every **6** hours

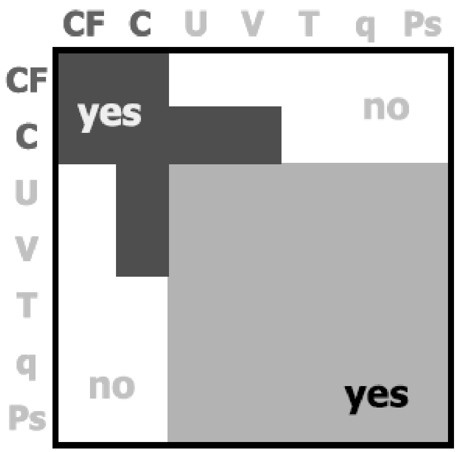
"Localization of variables" (Kang et al, JGR 2011)

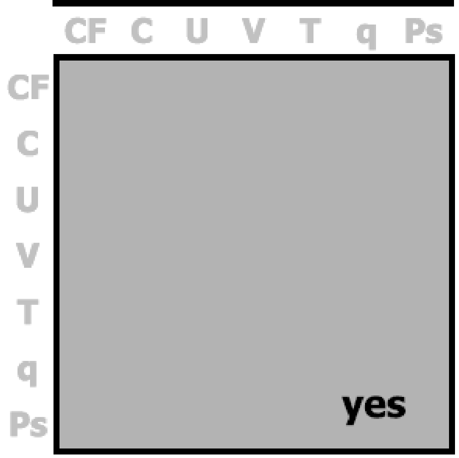


Schematic background error covariance matrix P^b.

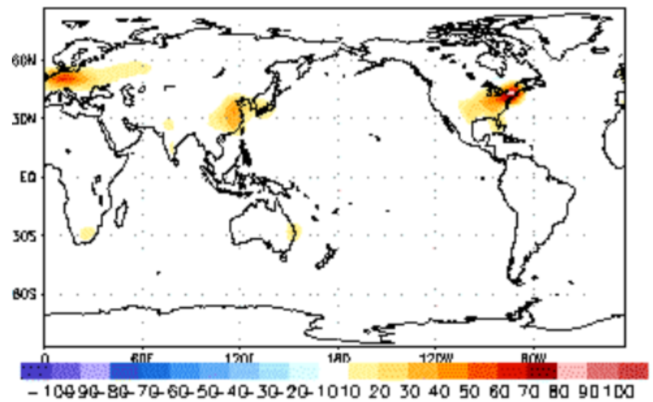
➔ **Zeroing out the background error covariance between unrelated variables improves the result of the analysis by reducing sampling errors.**

Results: Variable localization reduces sampling errors

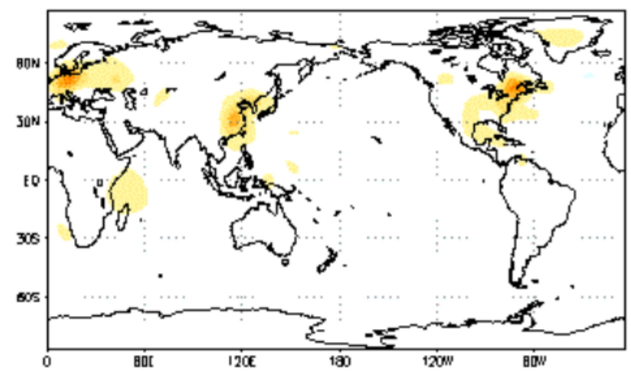




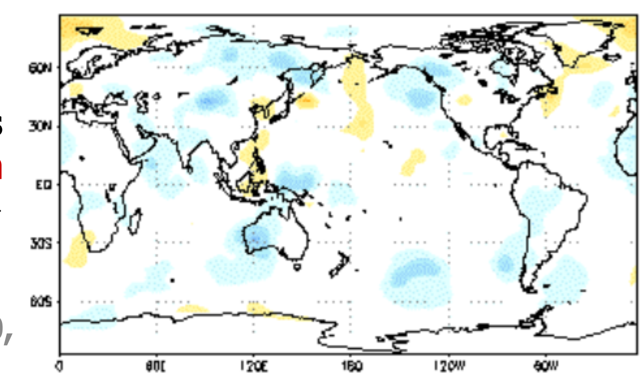
True CO_2 fluxes →
(anthropogenic)



Analysis of CO_2 fluxes
with variable localization
→



Analysis of CO_2 fluxes
without variable localization
→

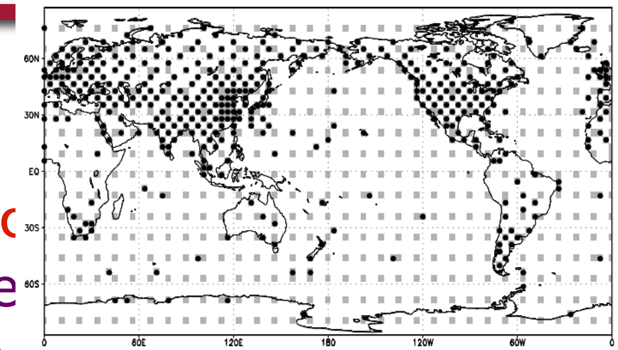


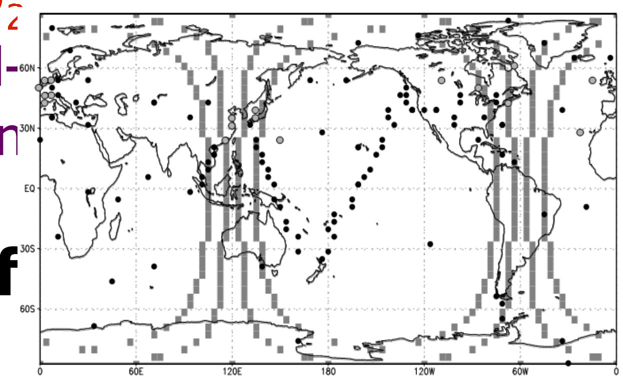
LETKF-C with SPEEDY-C

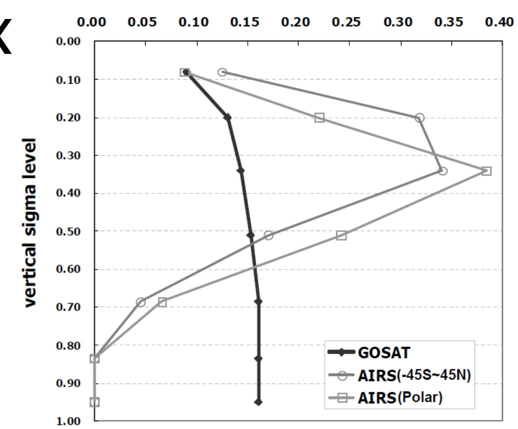
- Model: **SPEEDY-C** (Molteni, 2003; Kang, 2009)
  - Spectral AGCM model with T30L7
  - Prognostic variables: U, V, T, q, Ps, C
    - C (atmospheric CO_2): an inert tracer
  - **Persistence** forecast of Carbon Fluxes (CF), no observations
- **True CO2 fluxes: From CASA (Gurney et al, 2004)**
- Simulated observations
  - Rawinsonde observations of U, V, T, q, Ps
  - Ground-based observations of atmospheric CO_2
    - 18 hourly and 107 weekly data on the globe
  - Remote sensing data of column mixing CO_2
    - **AIRS** whose averaging kernel peaks at mid-troposphere
    - **GOSAT** whose averaging kernel is nearly uniform throughout the column
- Initial condition: random (**no *a-priori* information**)
- 20 ensembles

LETKF-C with SPEEDY-C

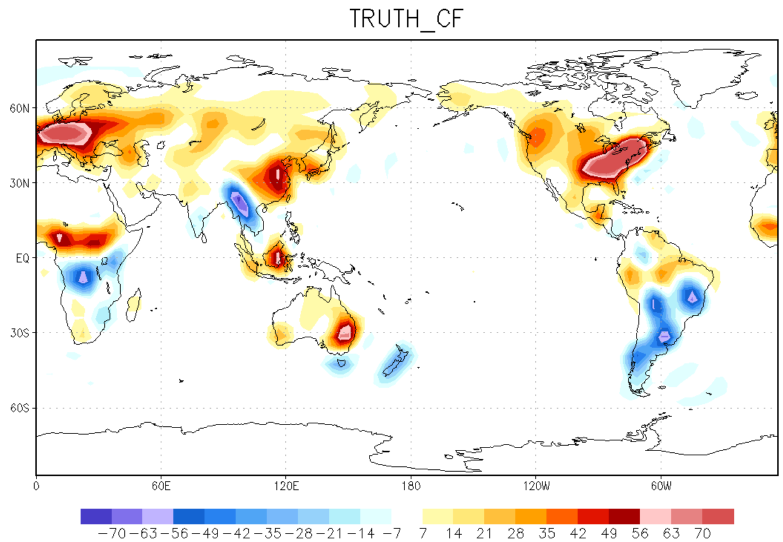
- Simulated observations
  - Rawinsonde observations of U, V, T, q, Ps
  - Ground-based observations of atmospheric
    - 18 hourly and 107 weekly data on the globe
  - Remote sensing data of column mixing CO_2
    - **AIRS** whose averaging kernel peaks at mid-
    - **GOSAT** whose averaging kernel is nearly un column
- Initial conditions: random (**no *a-priori* inf**
- 20 ensemble members
- No direct measurement of surface Carbon Flux
- CF only changes through the LETKF:
  persistence forecast.

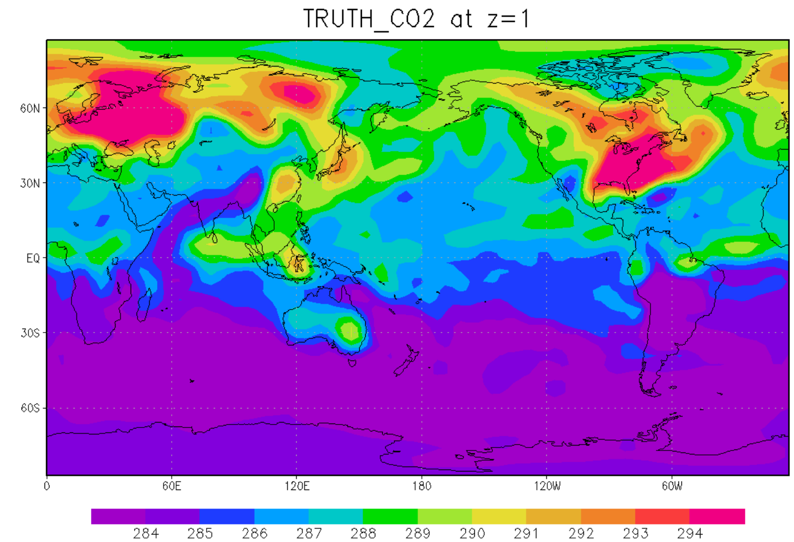


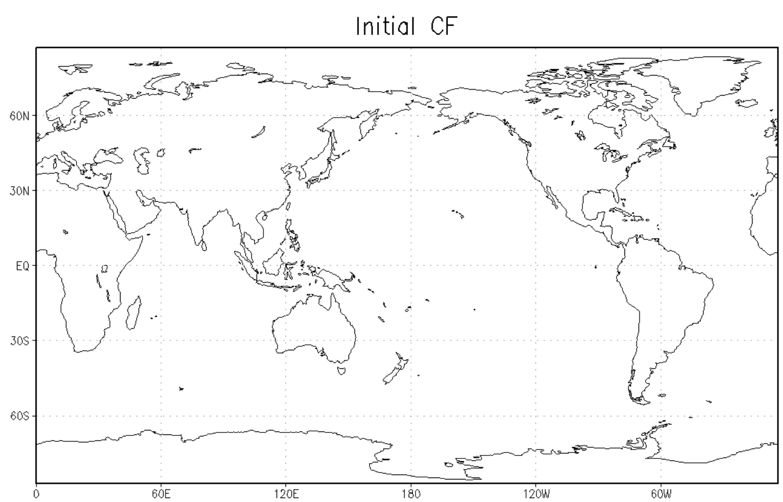


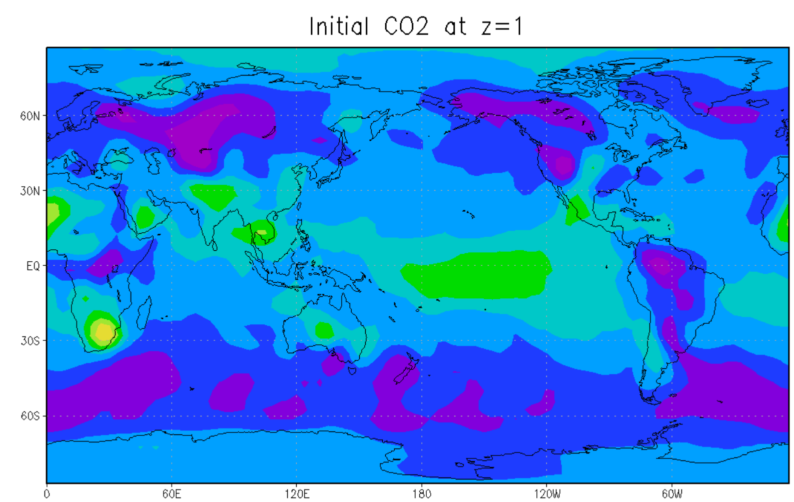


Initial conditions: random, no a priori information



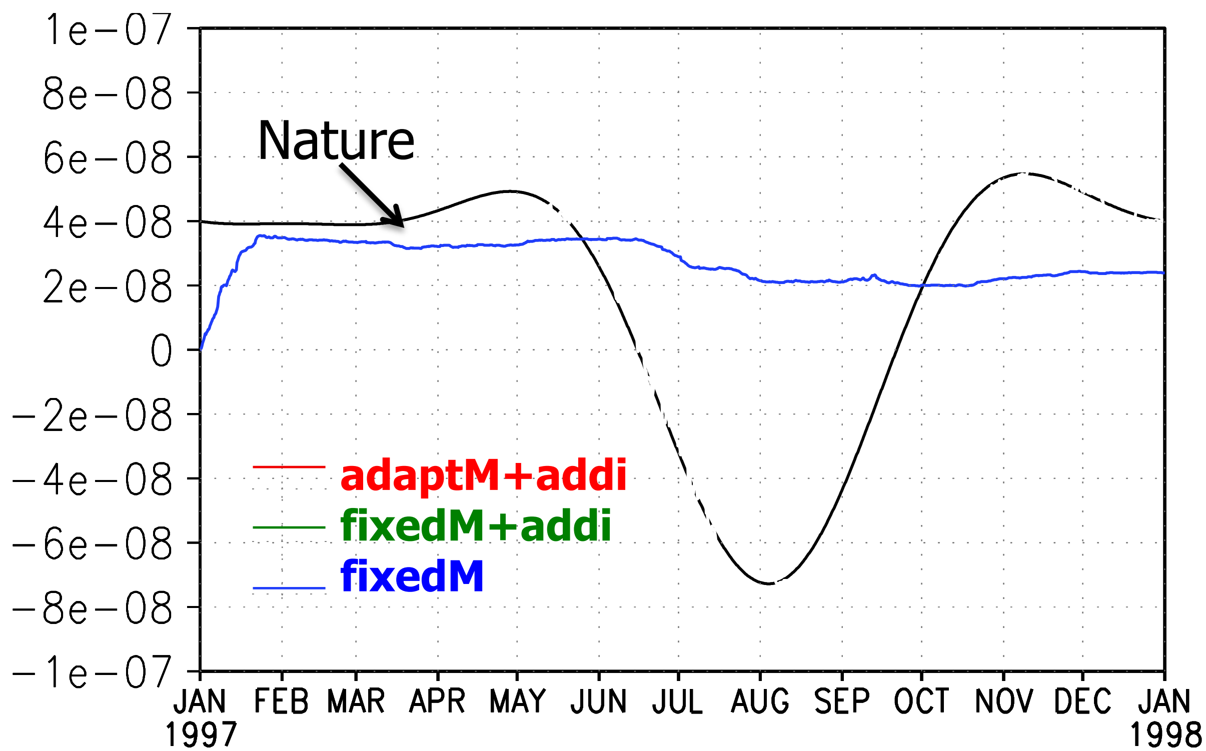




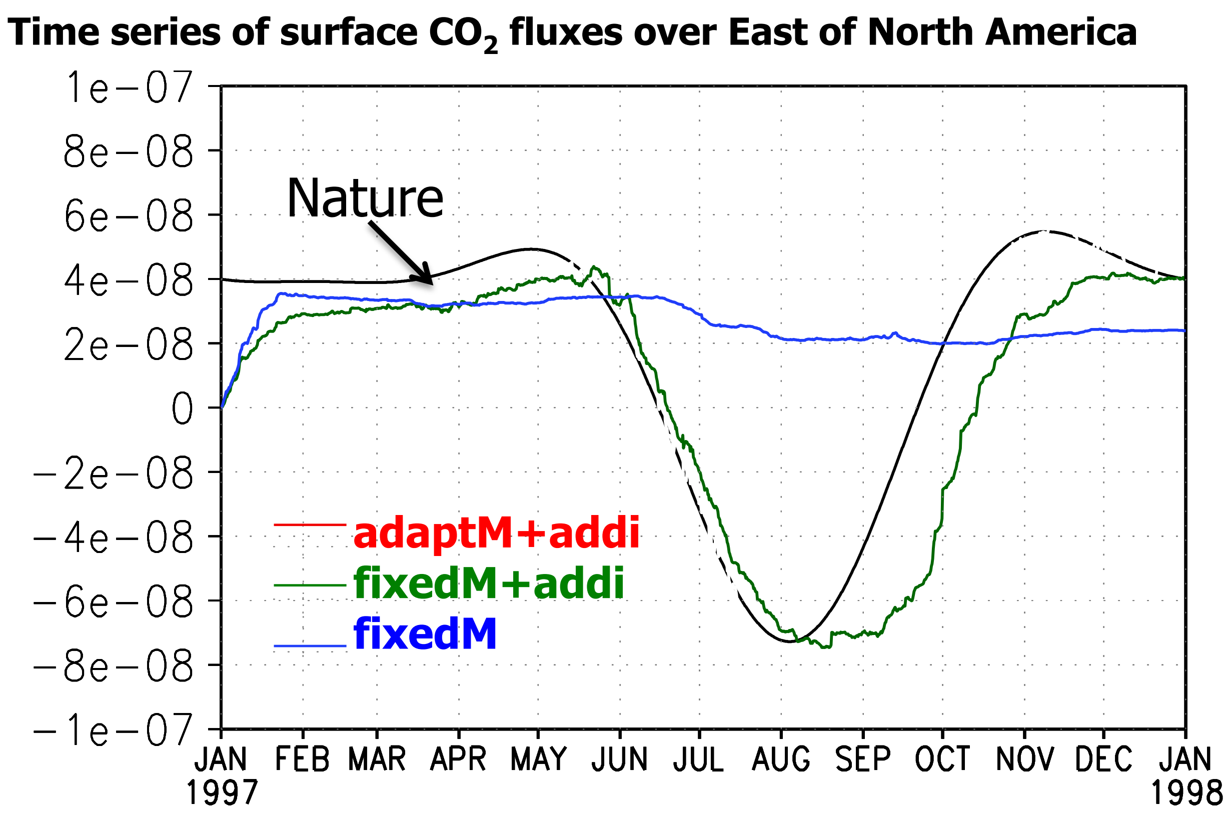


Impact of inflation: fixed multiplicative

Time series of surface CO_2 fluxes over East of North America

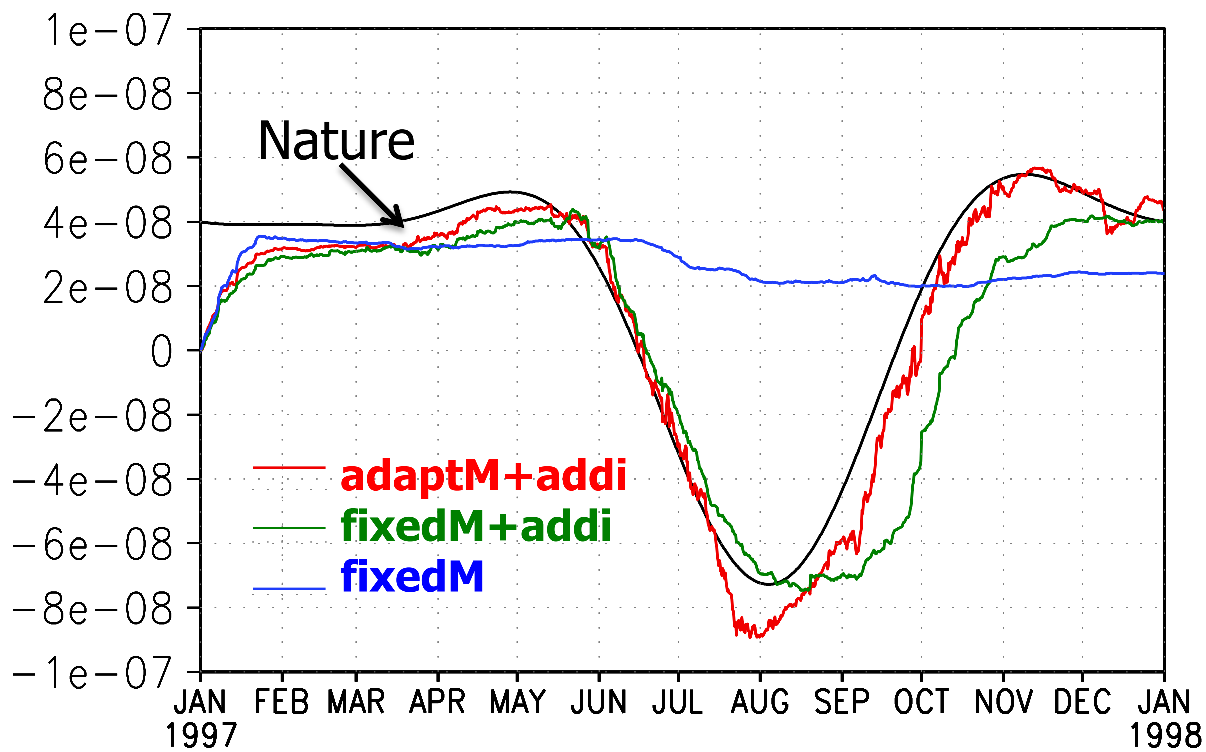


Impact of inflation: fixed multiplicative+additive



Impact of inflation: Adaptive multiplicative+additive

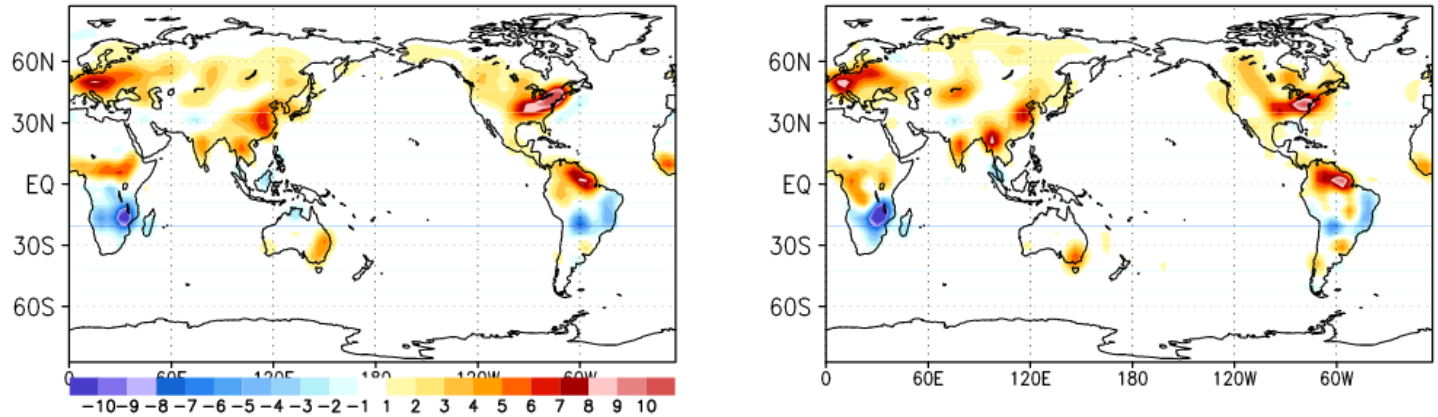
Time series of surface CO_2 fluxes over East of North America



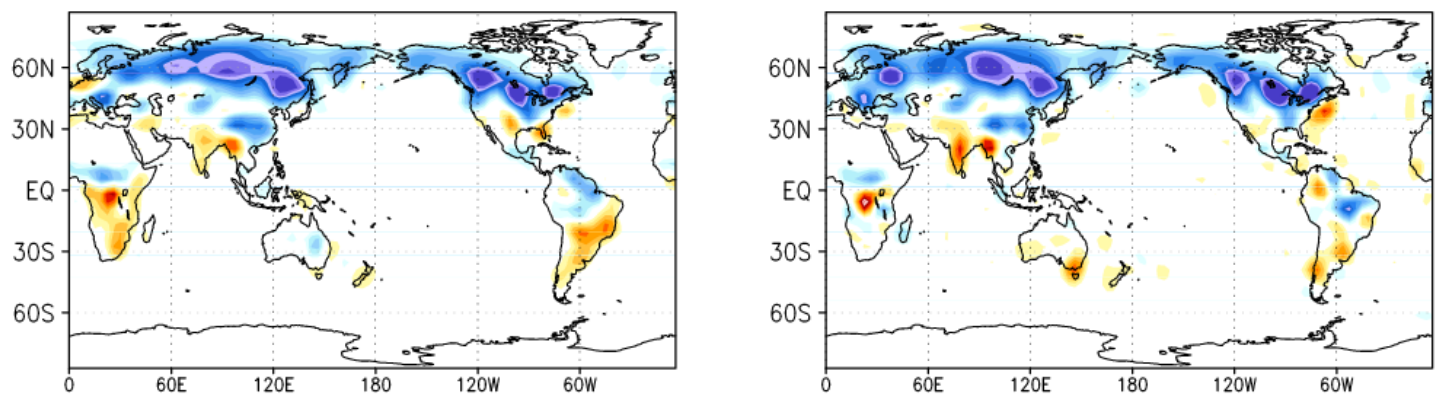
Results

▼ True CF ▼ ▼ Analysis of CF ▼

00Z01APR ▶
After three months of DA

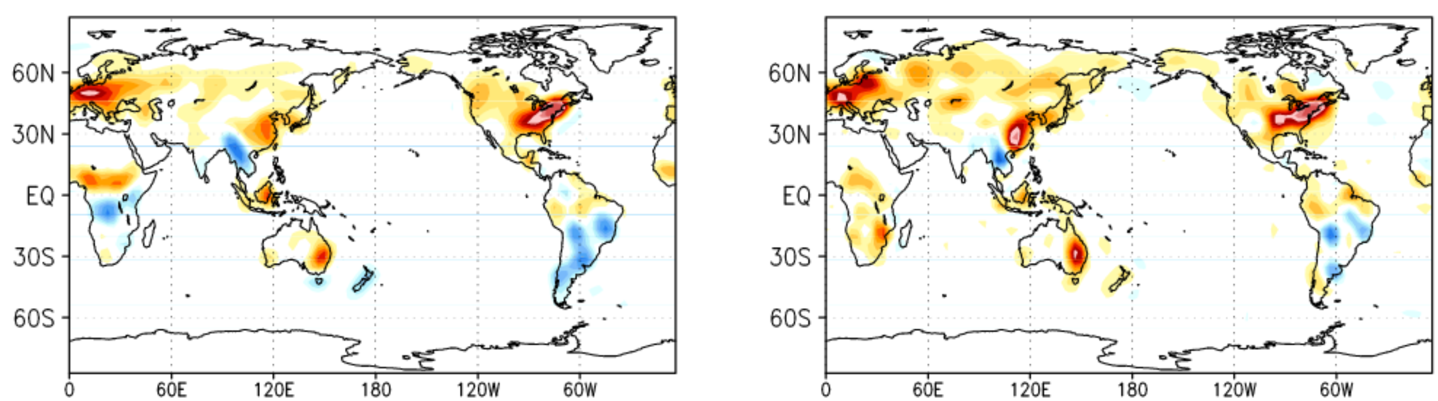


00Z01AUG ▶
After seven months of DA

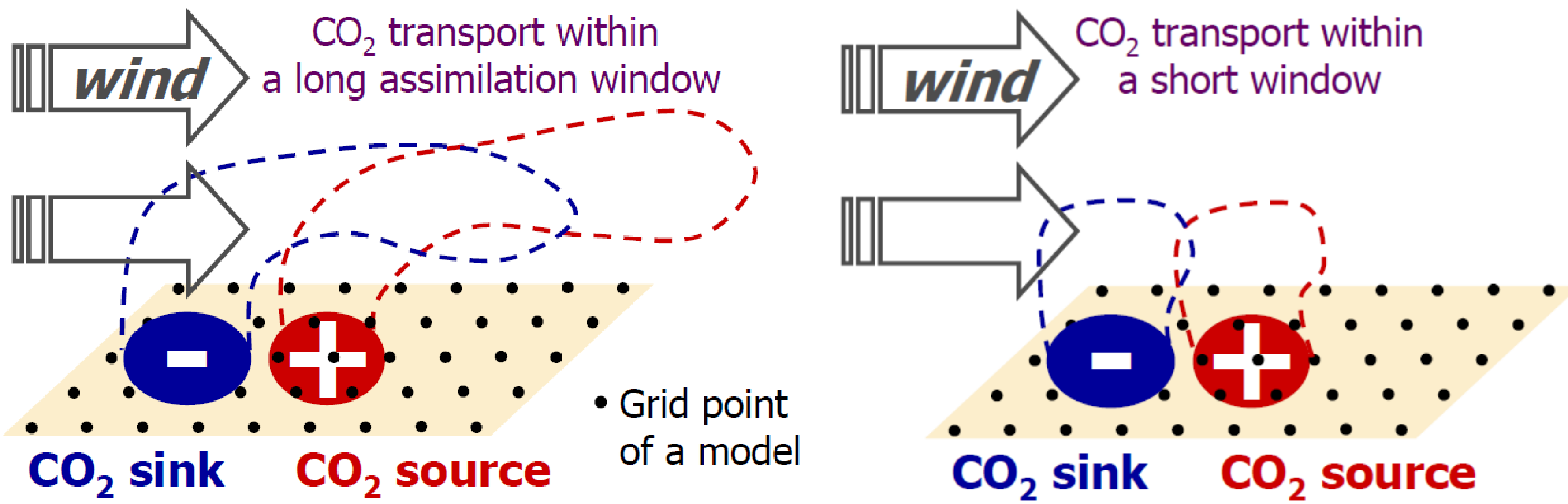


We succeeded in estimating time-evolving CF at model-grid scale!

00Z01JAN ▶
After one year of DA



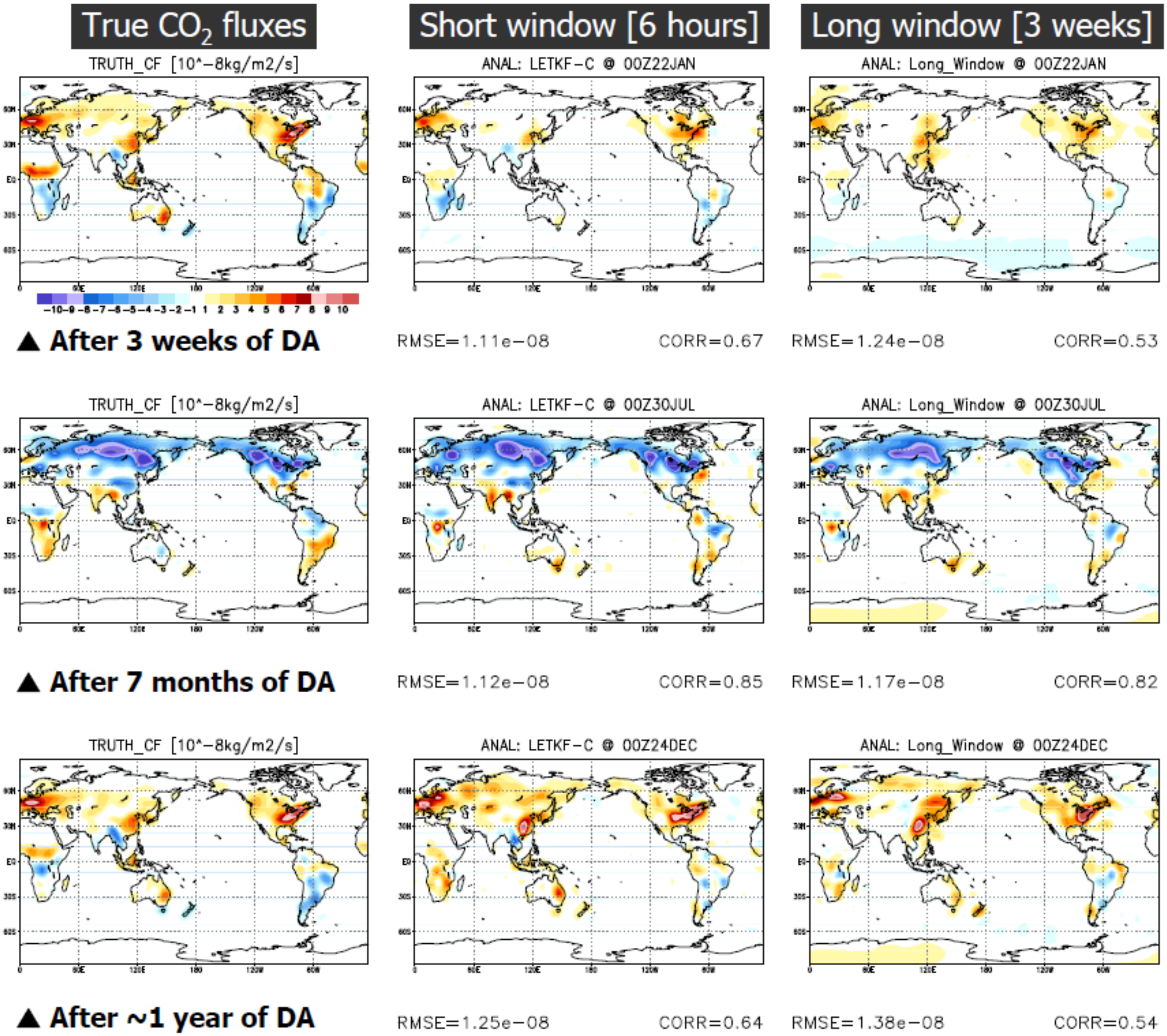
Assimilation window for Carbon fluxes inversion: current systems use a very long window



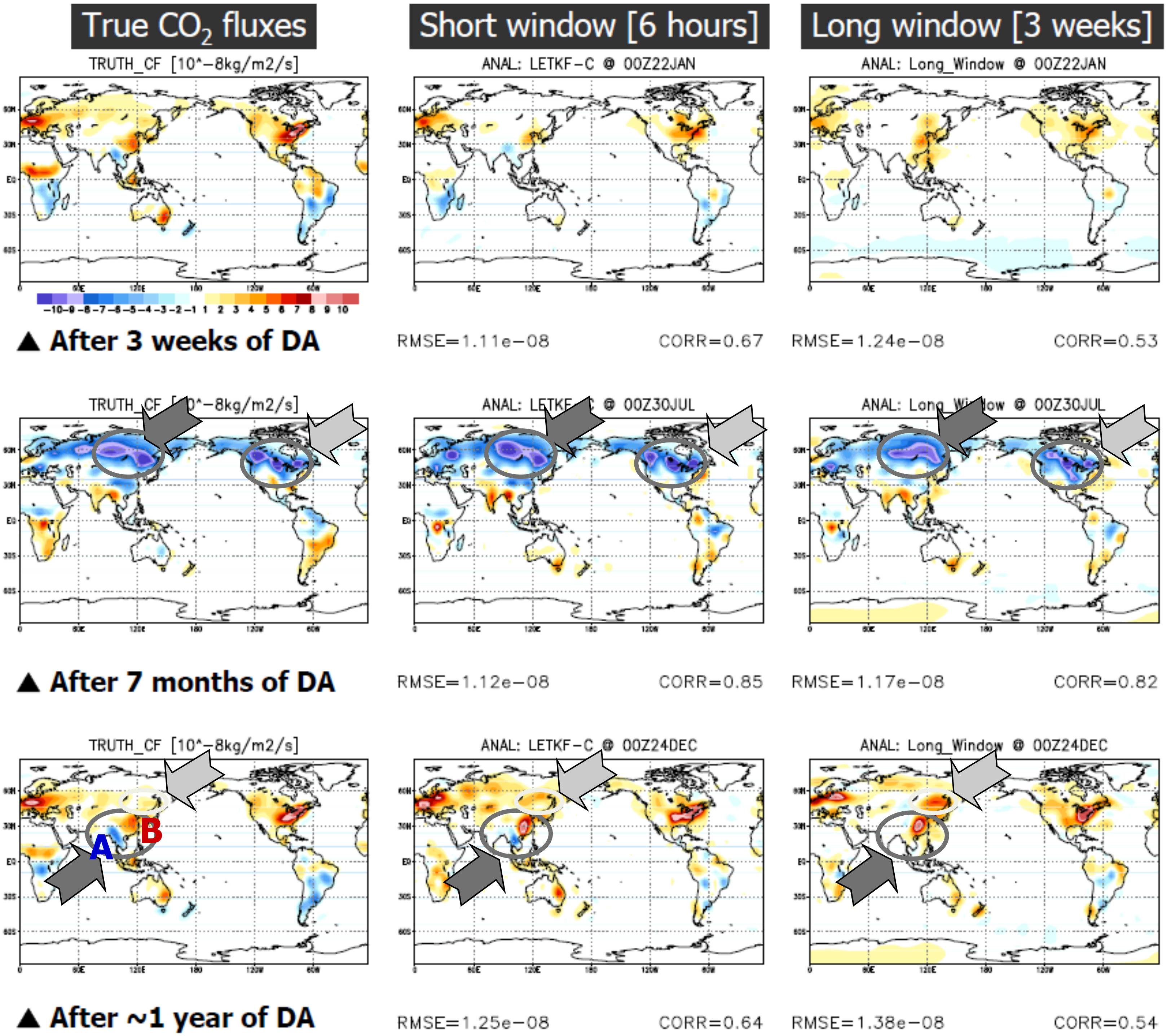
- CO_2 data assimilation system
  - A short assimilation window reduces the attenuation of observed CO_2 information because the analysis system can use the strong correlation between C and CF before the transport of atmospheric CO_2 blurs out the essential information of surface CO_2 forcing
  - We may not be able to reflect the optimal correlation between C and CF within a long assimilation window, which can introduce sampling errors into the EnKF analysis

Long vs. short window in LETKF-C

- **OSSEs with SPEEDY-C**
  - Realistic observation distributions for meteorological variables and CO_2
    - Rawinsonde observation for (U, V, T, q, Ps)
    - Ground-based observations, AIRS and GOSAT CO_2 mixing ratio for C
- **Experiment 1: Analysis from LETKF-C**
  - Simultaneous analysis with a 6-hour assimilation window
- **Experiment 2: Analysis from a long (3-week) assimilation window**
  - With this long assimilation window, **ensemble perturbations** of meteorological variables become **non-linear** so that we do not include wind uncertainty for CO_2 data assimilation (**Carbon-Univariate DA**)



True CO2 fluxes
Short window [6 hours]
Long window [3 weeks]
TRUTH_CF [10^-8kg/m2/s]
ANAL: LETKF-C @ 00Z22JAN
ANAL: Long_Window @ 00Z22JAN
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10
▲ After 3 weeks of DA
RMSE=1.11e-08
CORR=0.67
RMSE=1.24e-08
CORR=0.53
TRUTH_CF [10^-8kg/m2/s]
ANAL: LETKF-C @ 00Z30JUL
ANAL: Long_Window @ 00Z30JUL
▲ After 7 months of DA
RMSE=1.12e-08
CORR=0.85
RMSE=1.17e-08
CORR=0.82
TRUTH_CF [10^-8kg/m2/s]
ANAL: LETKF-C @ 00Z24DEC
ANAL: Long_Window @ 00Z24DEC
▲ After ~1 year of DA
RMSE=1.25e-08
CORR=0.64
RMSE=1.38e-08
CORR=0.54

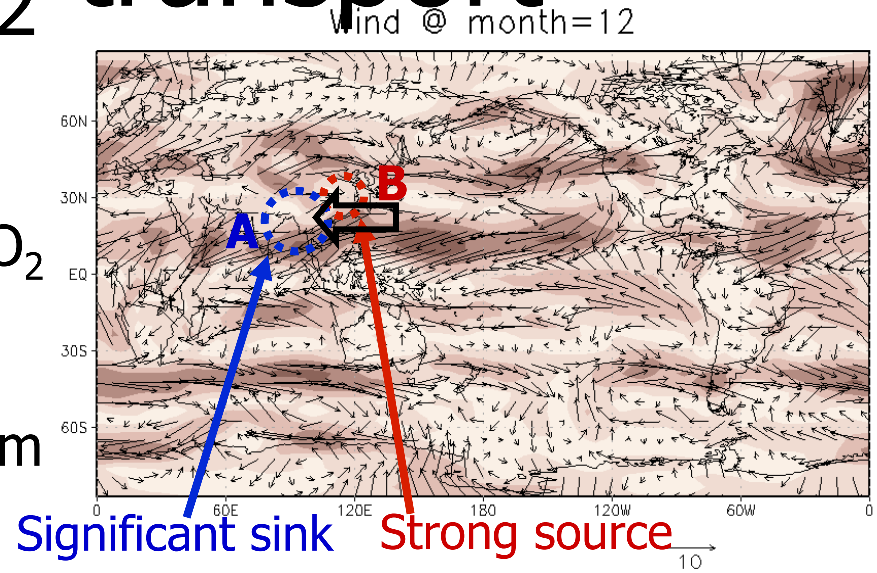


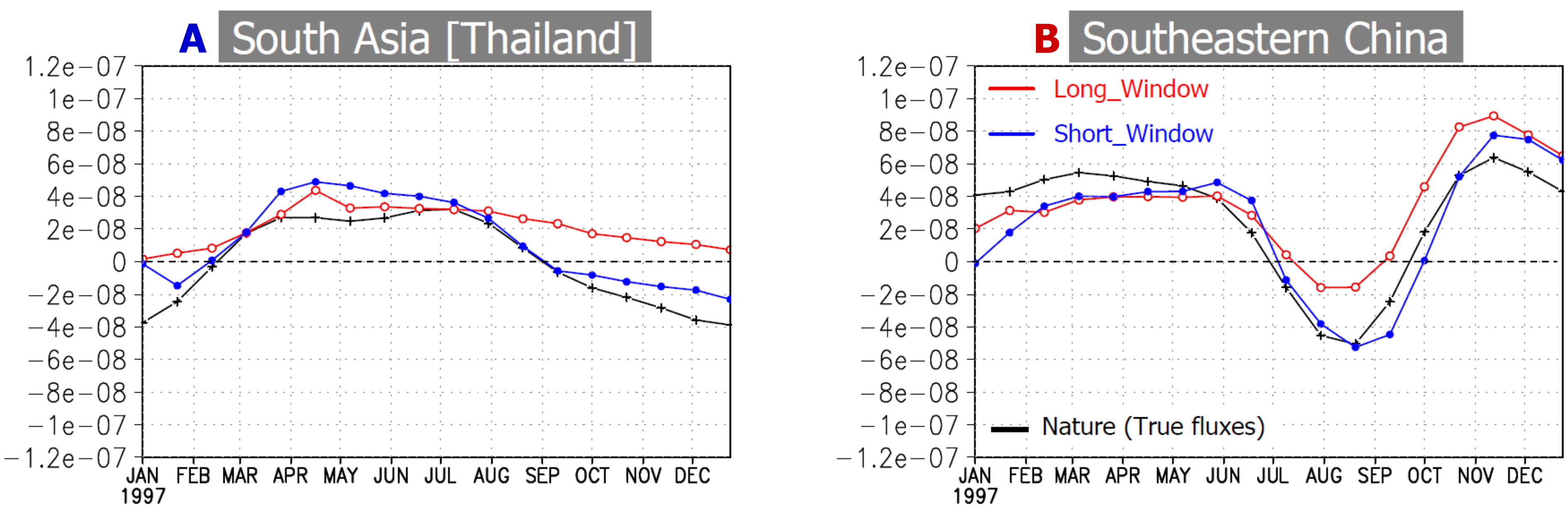
True CO2 fluxes
Short window [6 hours]
Long window [3 weeks]
TRUTH_CF [10^-8kg/m2/s]
ANAL: LETKF-C @ 00Z22JAN
ANAL: Long_Window @ 00Z22JAN
▲ After 3 weeks of DA
RMSE=1.11e-08
CORR=0.67
RMSE=1.24e-08
CORR=0.53
ANAL: LETKF-C @ 00Z30JUL
ANAL: Long_Window @ 00Z30JUL
▲ After 7 months of DA
RMSE=1.12e-08
CORR=0.85
RMSE=1.17e-08
CORR=0.82
TRUTH_CF [10^-8kg/m2/s]
A
B
ANAL: LETKF-C @ 00Z24DEC
ANAL: Long_Window @ 00Z24DEC
▲ After ~1 year of DA
RMSE=1.25e-08
CORR=0.64
RMSE=1.38e-08
CORR=0.54

Impact of CO_2 transport

- **Strong easterly** from the source region to the sink region brings CO_2 increase information over the sink area

➔ There are incorrect positive CF from OCT to DEC (the end of DA)





Time series of surface CO_2 flux estimations

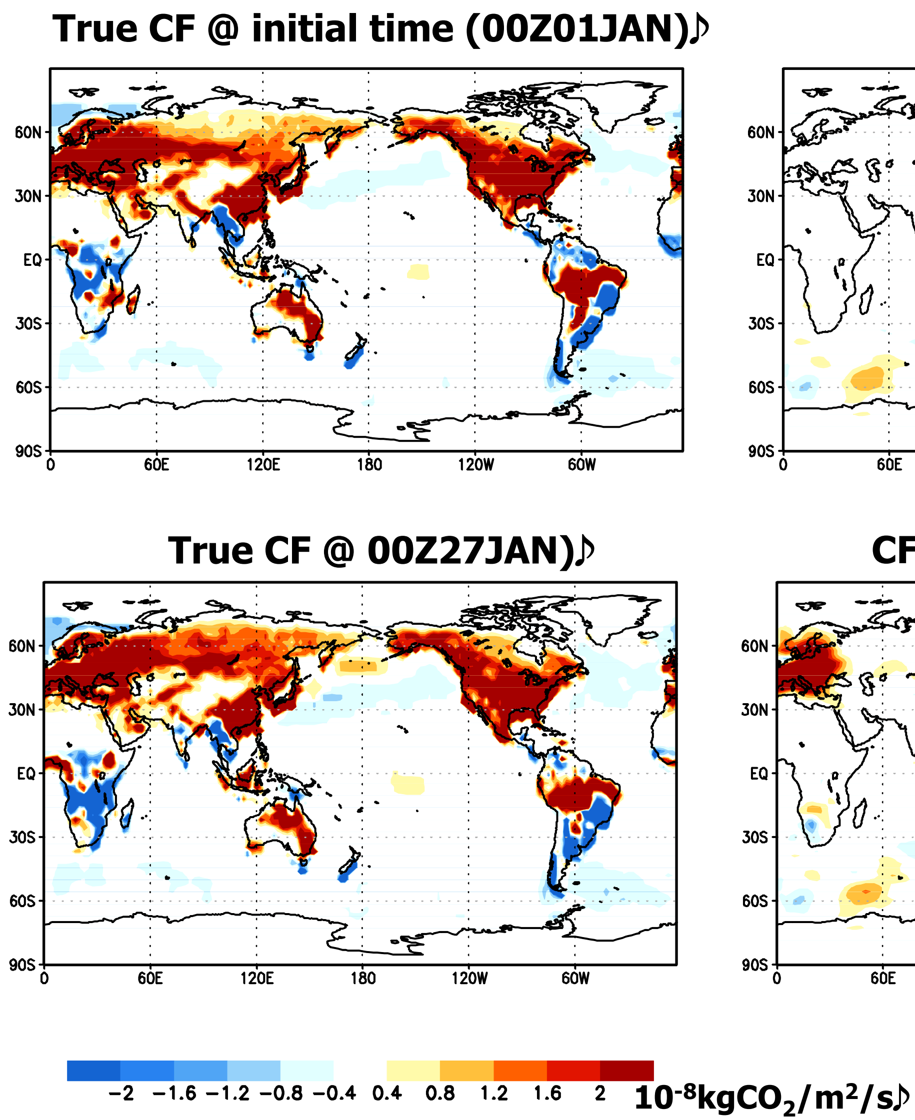
Summary of LETKF-C carbon fluxes

- **Assimilation window**
  - EnKF has better performance with a short window
  - CO_2 observations may be able to provide some information to distant CF, but it becomes blurred (an ill-posed problem).
- Implement LETKF-C on the NCAR CAM model
  - OSSE with realistic observations
  - Very slow (only 26 days)
  - Preliminary results are encouraging

LETKF-C with NCAR CAM3.5

- Model: **CAM 3.5**
  - Finite Volume dynamical core
  - $2.5^\circ \times 1.9^\circ$ of horizontal resolution with 26 layers in the vertical
  - C (atmospheric CO_2) is an inert tracer
  - Persistence forecast of CF
- Simulated observations with **real observation coverage**
  - Conventional data for U, V, T, q, Ps
  - Ground-based observations of atmospheric CO_2
    - ~10 hourly and ~100 weekly records on the globe
  - Remote sensing data of column mixing CO_2
    - **AIRS** whose averaging kernel peaks at mid-troposphere
- Initial conditions: random (no *a-priori* information)
- 64 ensembles

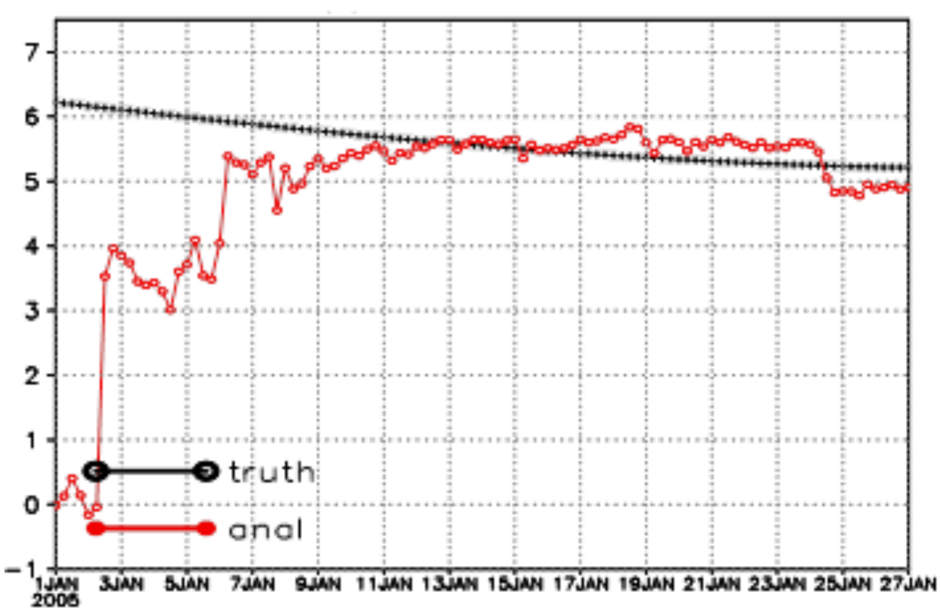
LETKF-CAM 3.5 analysis



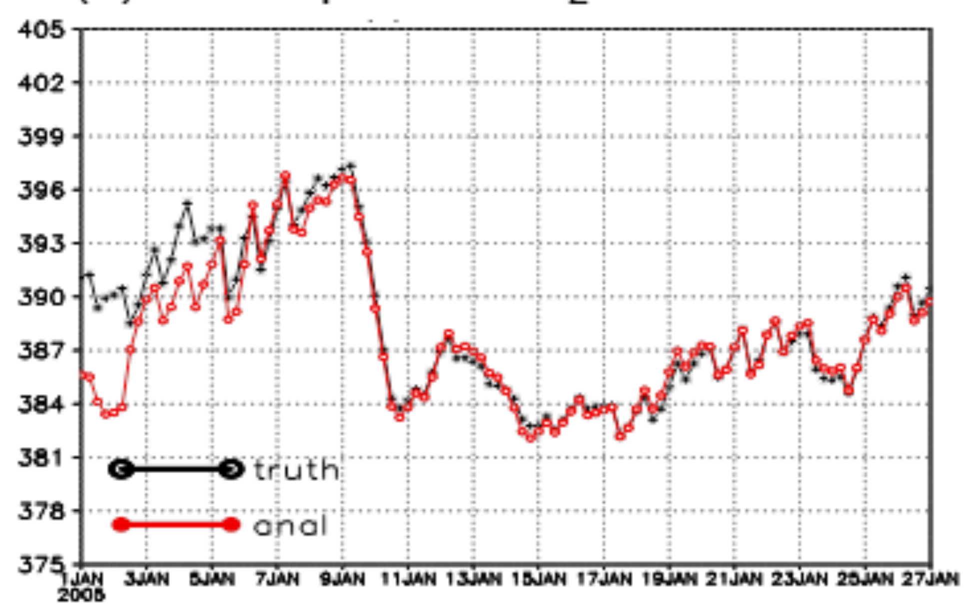
LETKF-CAM3.5 CF analysis♪

- Time series of surface CO_2 fluxes and atmospheric CO_2 concentrations over Europe (observation-rich area)♪

(a) Surface CO_2 fluxes over EUR



(b) Atmospheric CO_2 over EUR

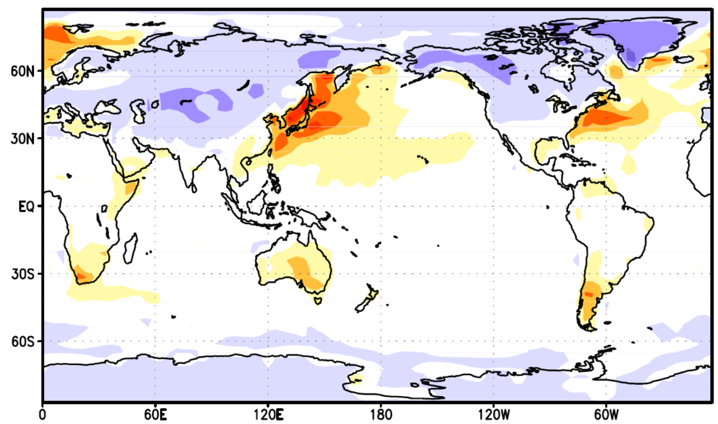


Surface Heat and Moisture Fluxes

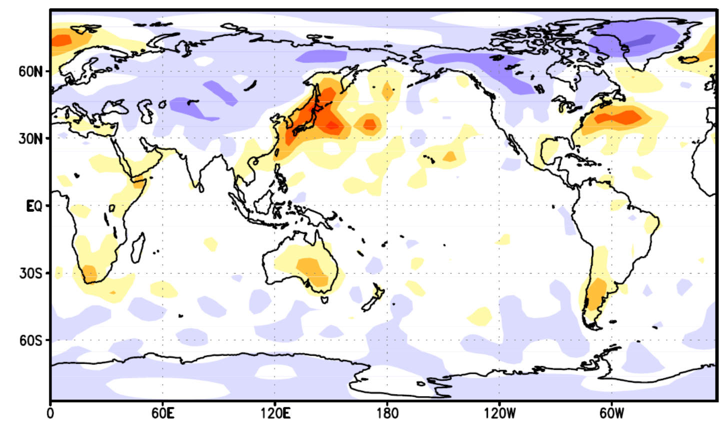
- Can we estimate **surface moisture/heat fluxes** by assimilating atmospheric moisture/temperature observations? *We can use the same methodology!*
- OSSEs
  - Nature: SPEEDY (perfect model)
  - Forecast model: SPEEDY with **persistence forecast of Sensible/Latent heat fluxes (SHF/LHF)**
  - Observations: conventional observations of (U, V, T, q, Ps) and AIRS retrievals of (T, q)
  - Analysis: U, V, T, q, Ps + SHF & LHF
- Fully multivariate data assimilation
- Adaptive multiplicative inflation + additive inflation
- Initial conditions: random (no *a-priori* information)

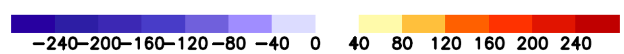
Results: SHF
(with perfect wind stress parameterization)

True SHF @ end of JAN

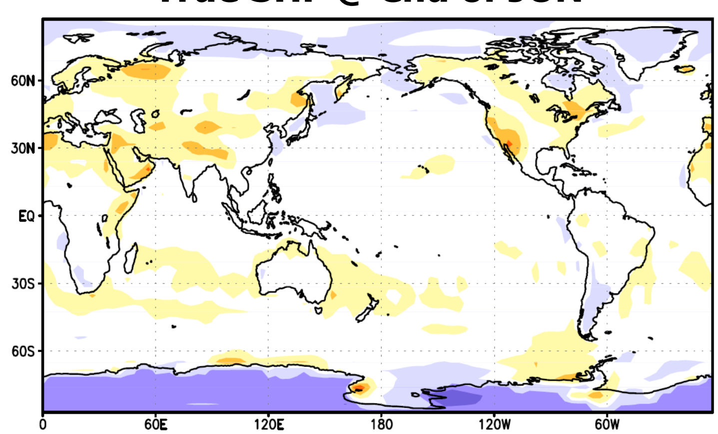


SHF analysis @ end of JAN

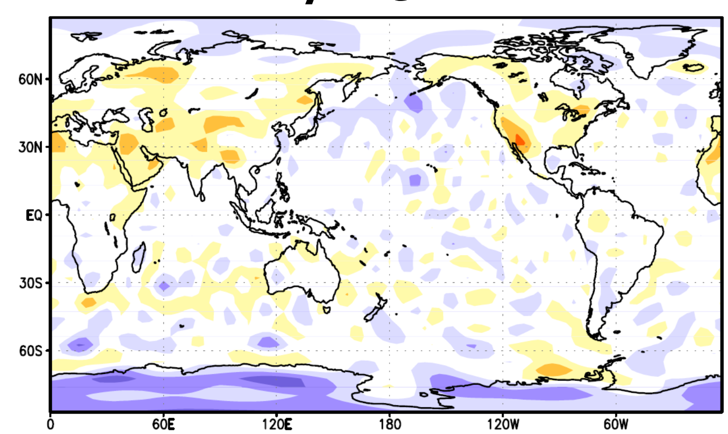




True SHF @ end of JUN

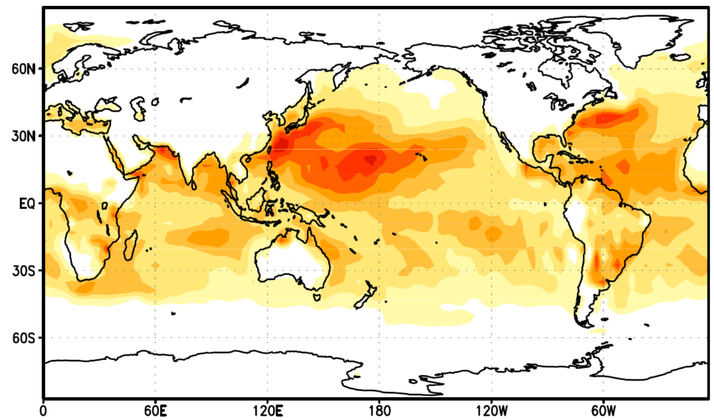


SHF analysis @ end of JUN

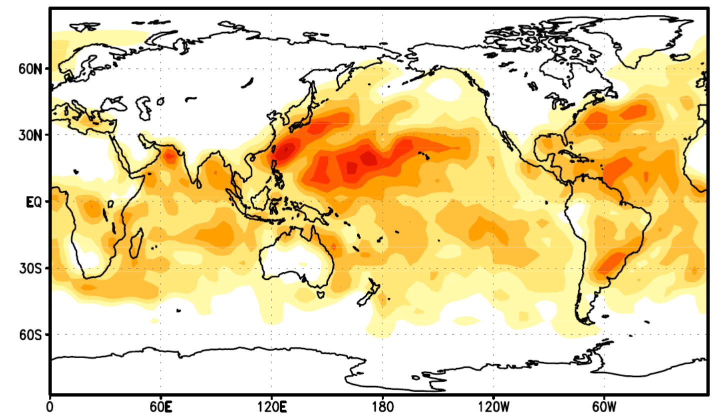


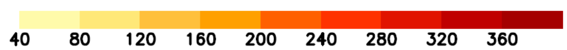
Results: LHF (perfect wind stress parameterization)

True LHF @ end of JAN

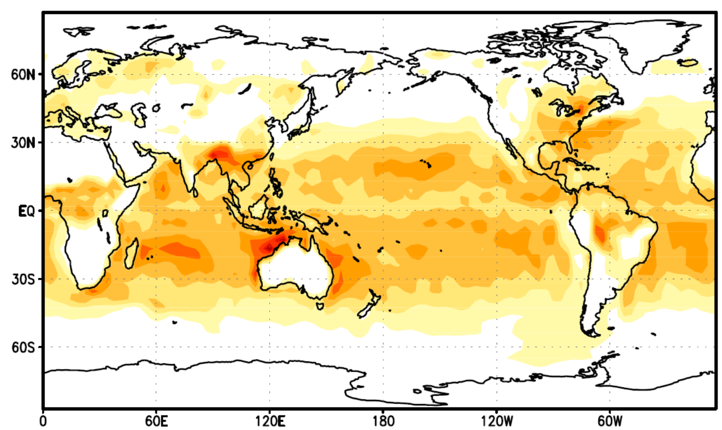


LHF analysis @ end of JAN

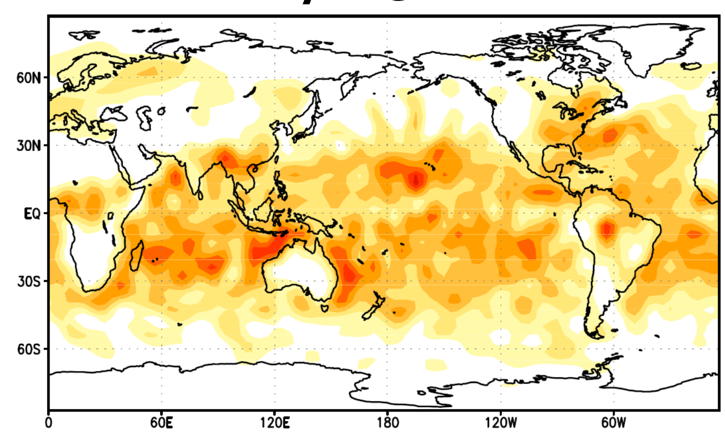




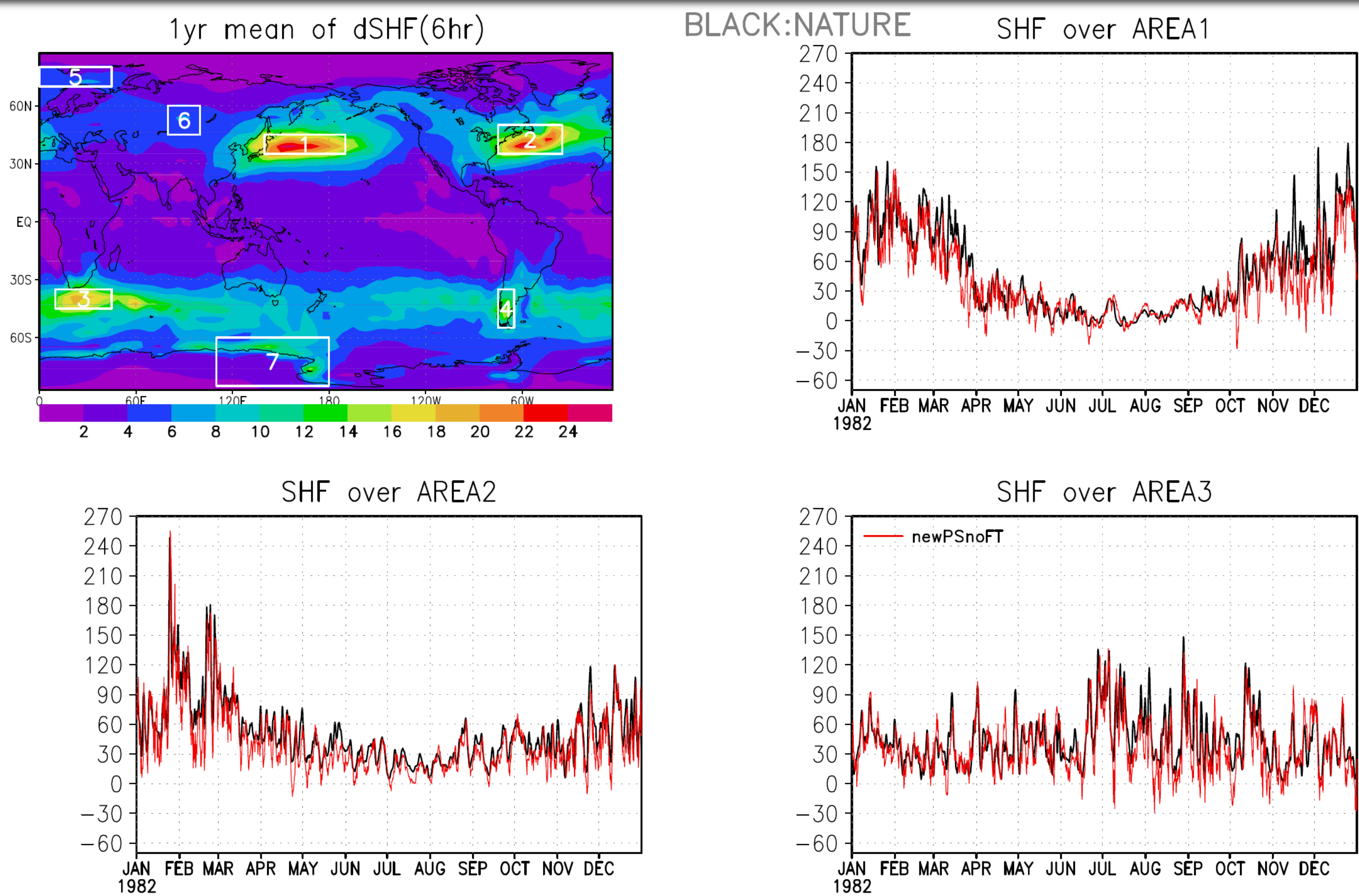
True LHF @ end of JUN



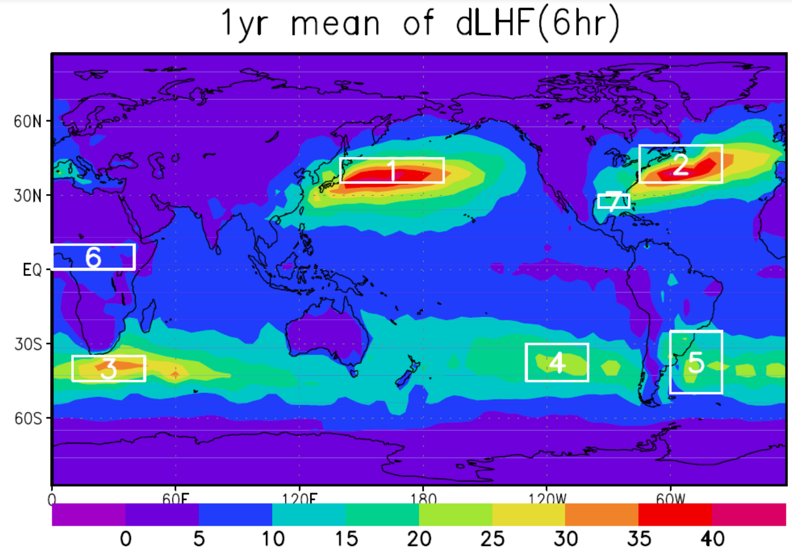
LHF analysis @ end of JUN

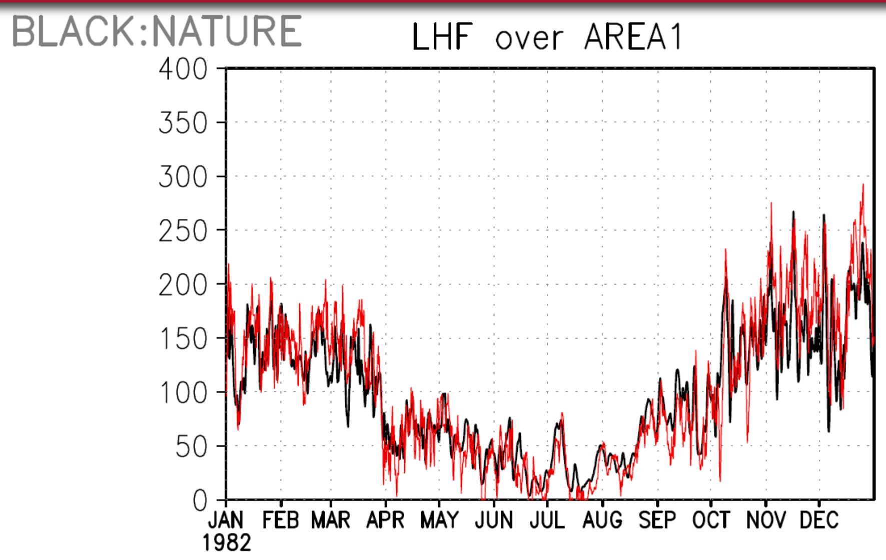


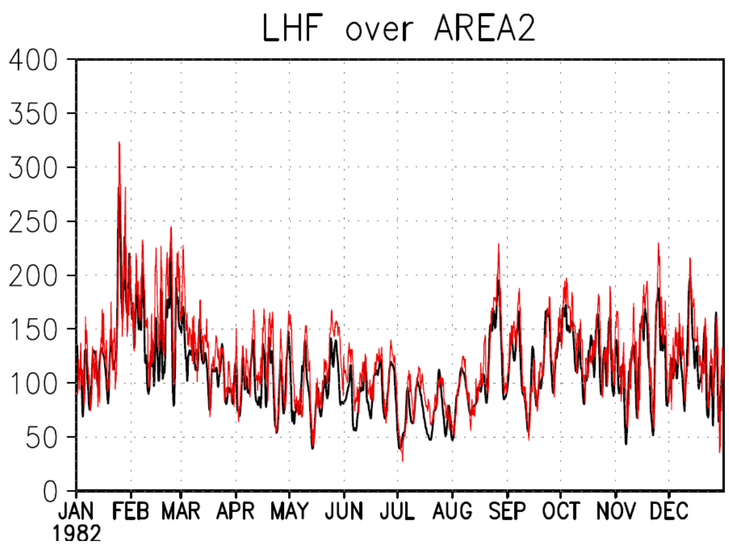
Time series of SHF (perfect wind stress parameterization)

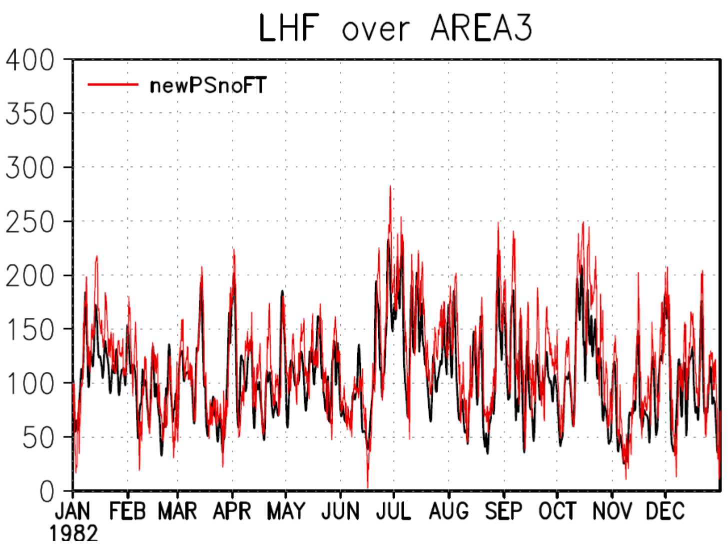


Time series of LHF (perfect wind stress parameterization)









Summary of SHF & LHF DA
(with perfect WSTR parameterization)

- AIRS retrievals of T and q provide accurate and abundant information for constraining surface heat and moisture fluxes
  - Observation error: 1K for T and 1.0g/kg for q
  - Global coverage at every 12 hours

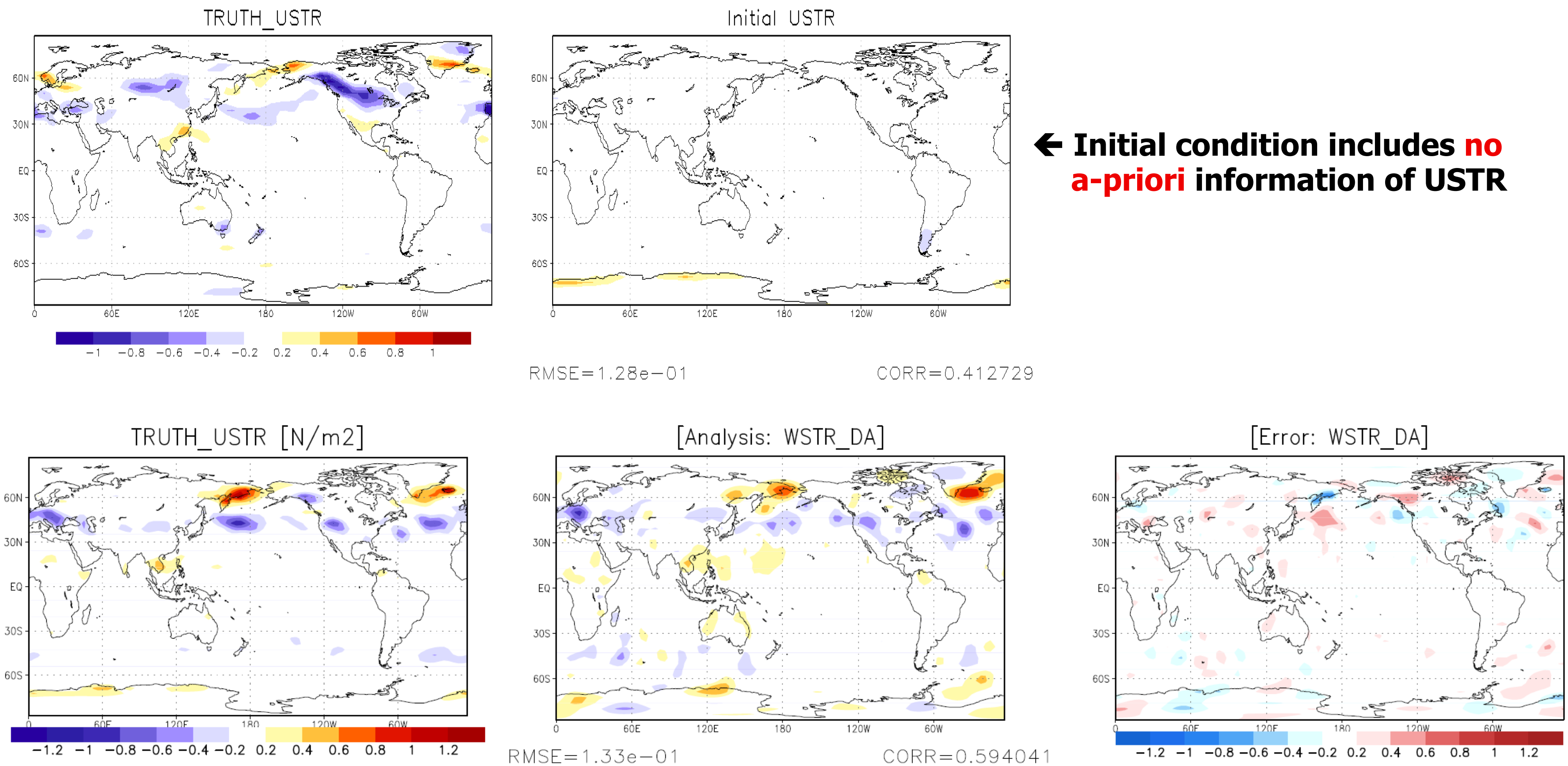
➔ After a short spin-up period (~a week), estimation of SHF and LHF converges very well

- But results shown here are given under the assumption of a perfect wind stress parameterization.

Can we also estimate wind stress?

- OSSEs
  - Nature: SPEEDY
  - Forecast model: SPEEDY with **persistence forecast of Sensible/Latent heat fluxes (SHF/LHF)** and wind stress **(USTR, VSTR)** [ALL_FLUXES]
  - Observations: conventional observations of (U, V, T, q, Ps), AIRS retrievals of (T, q), and ASCAT ocean surface wind observations
    - **Observation error of ASCAT: 3.5m/s (not as good as AIRS data)**
    - ASCAT covers the global ocean every 12 hours, but with little overlap with AIRS.
    - Analysis: U, V, T, q, Ps + SHF, LHF, USTR, VSTR
- Fully multivariate data assimilation
- Initial conditions: random (no *a-priori* information)

Result: USTR from [ALL_FLUXES]

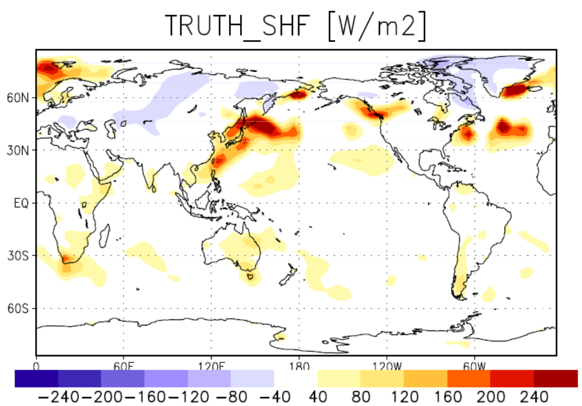


← **Initial condition includes no a-priori information of USTR**

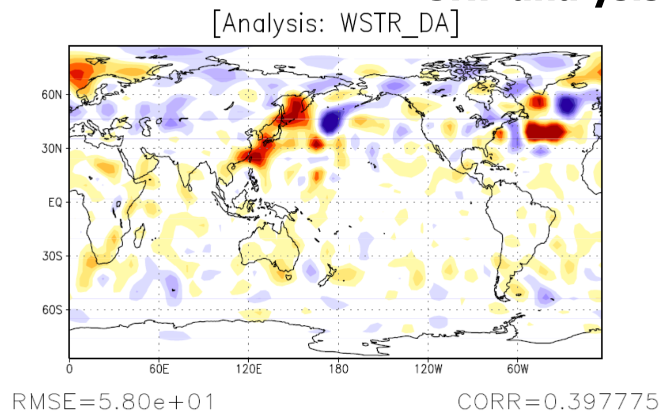
↑ **After one month of DA**, USTR estimation is close to the true USTR

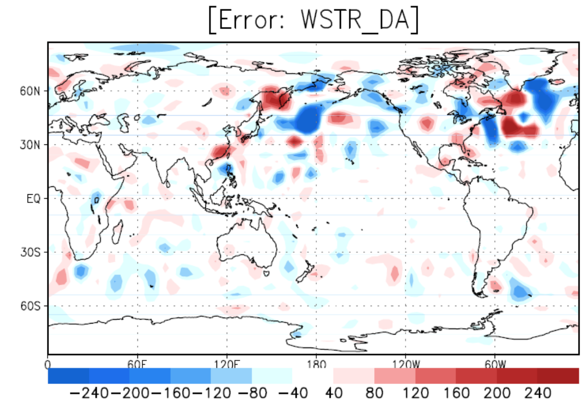
Results: SHF from [ALL_FLUXES]

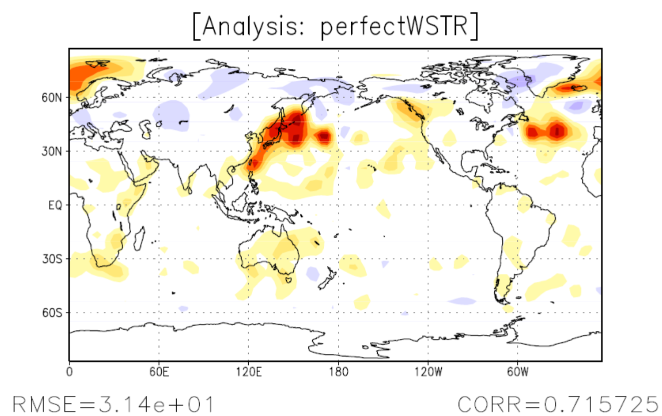
▼ SHF analysis with WSTR DA ▼

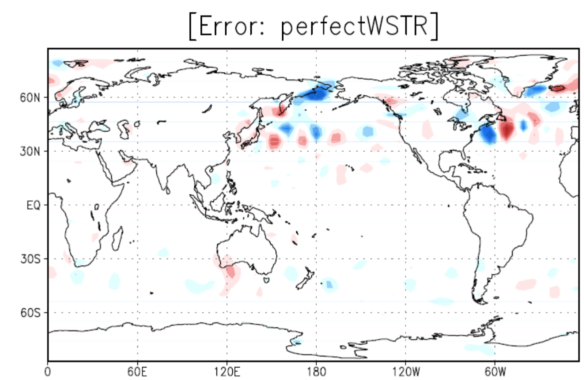


▲ True SHF







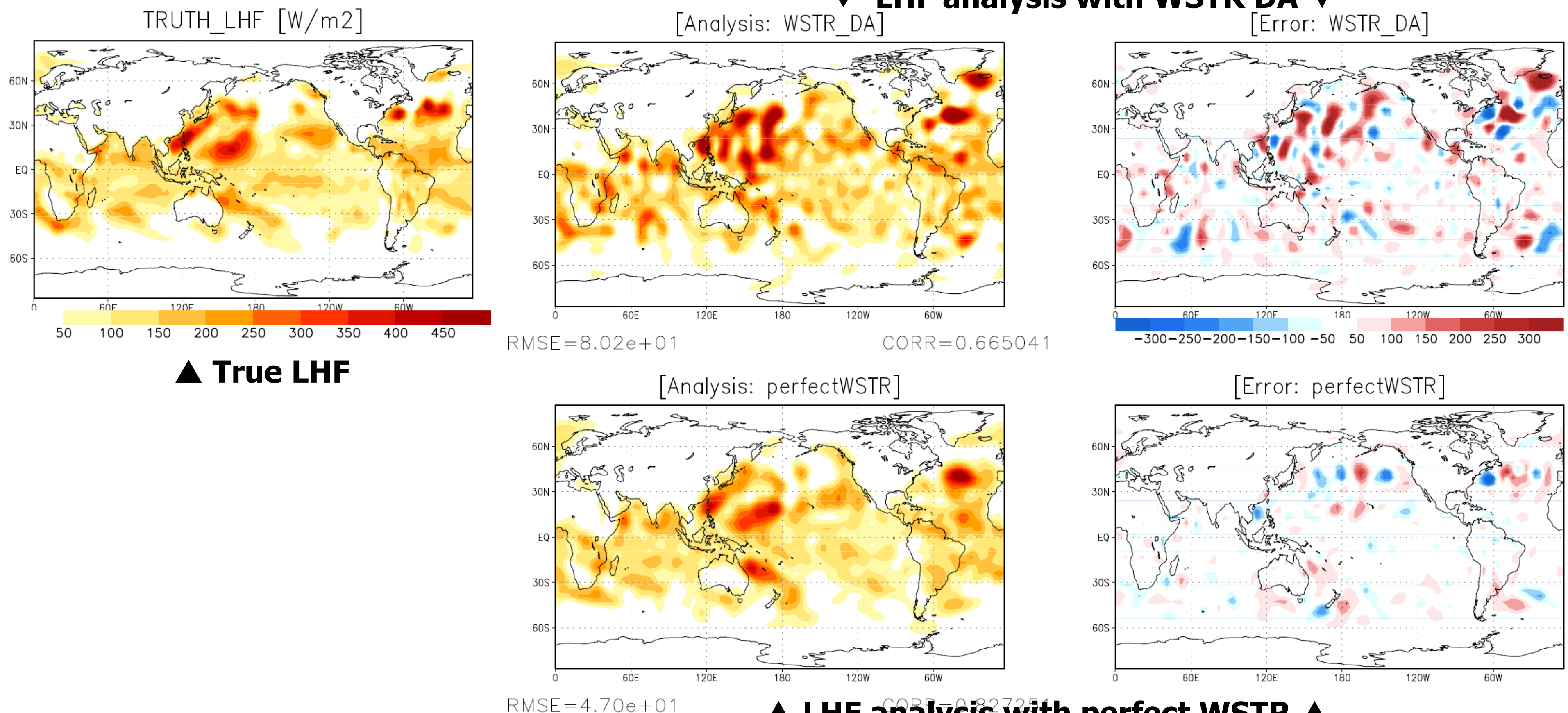


▲ SHF analysis with perfect WSTR ▲

- Although the estimated wind stress does look okay, the imperfection of the wind stress contaminates the estimation of SHF and LHF significantly

Results: LHF from [ALL_FLUXES]

▼ LHF analysis with WSTR DA ▼



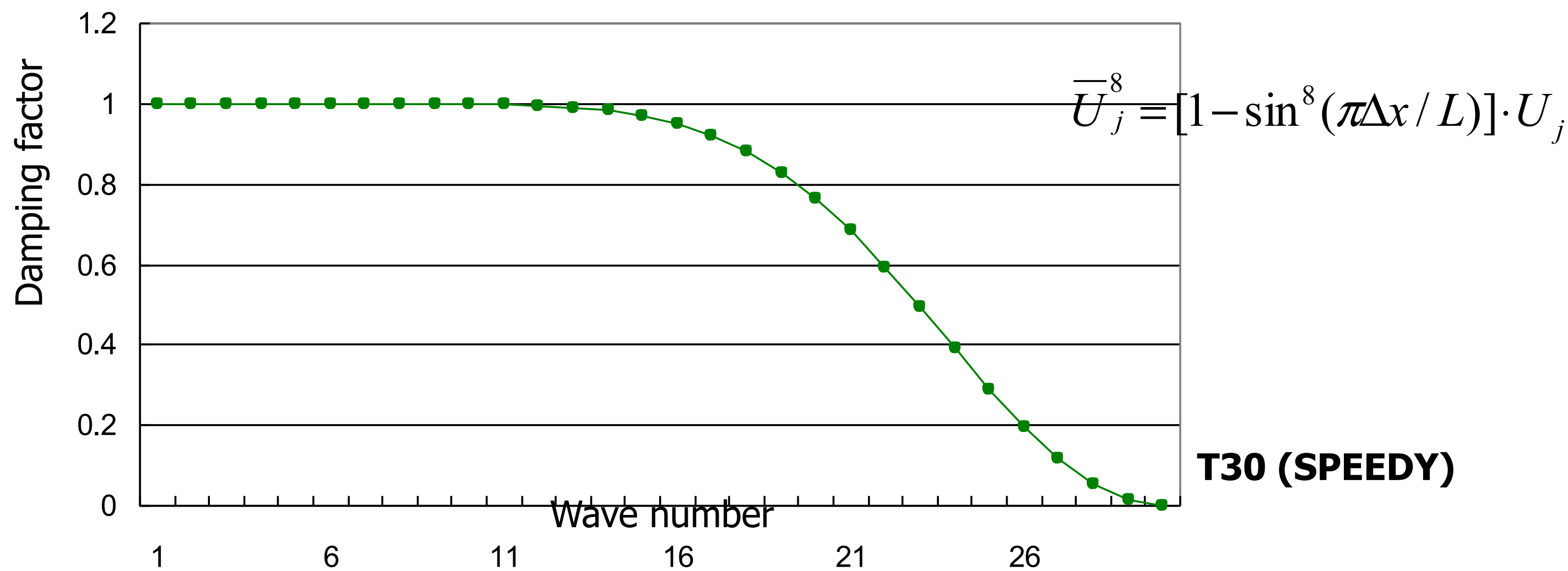
▲ LHF analysis with perfect WSTR ▲

- Although the estimated wind stress does look okay, the imperfection of the wind stress contaminates the estimation of SHF and LHF significantly

➔ ***Analyses diverged***…

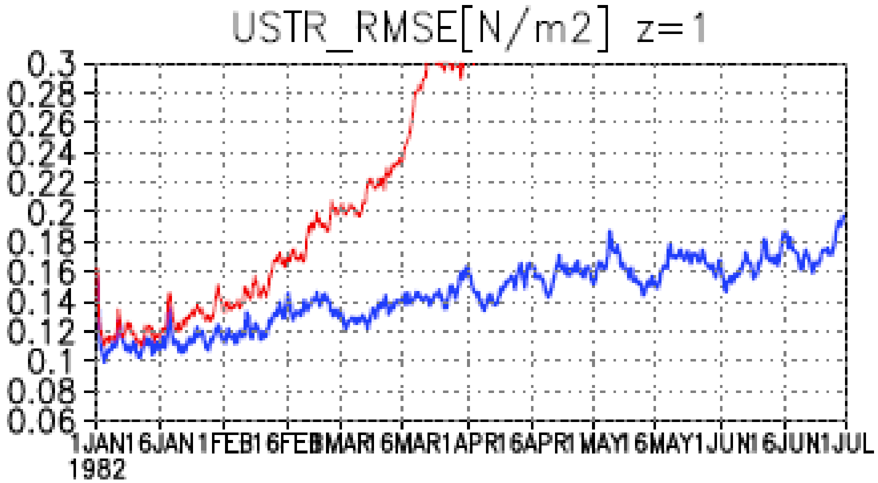
1) Filtering analysis increments?

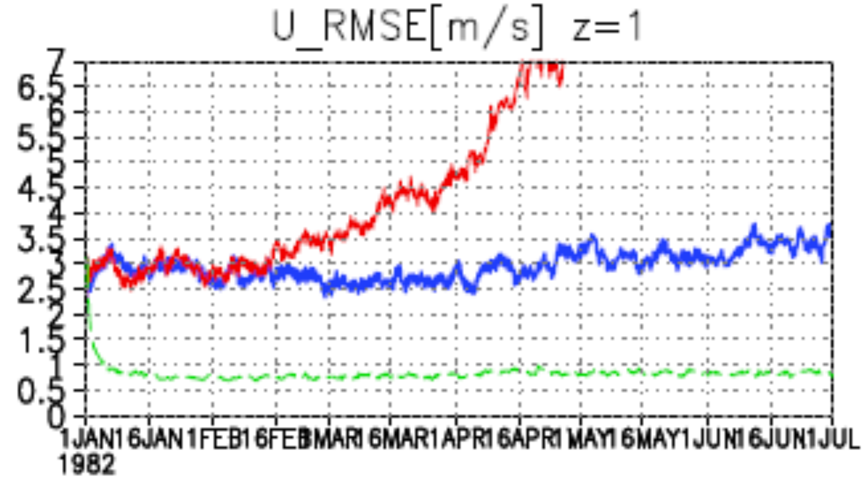
- Due to the **limited observational contents**, we may not be able to expect analysis increment with a full resolution
  - Filtering out high wavenumbers from the analysis increments for 2d parameters (SHF, LHF, USTR, VSTR) using the Shapiro filter

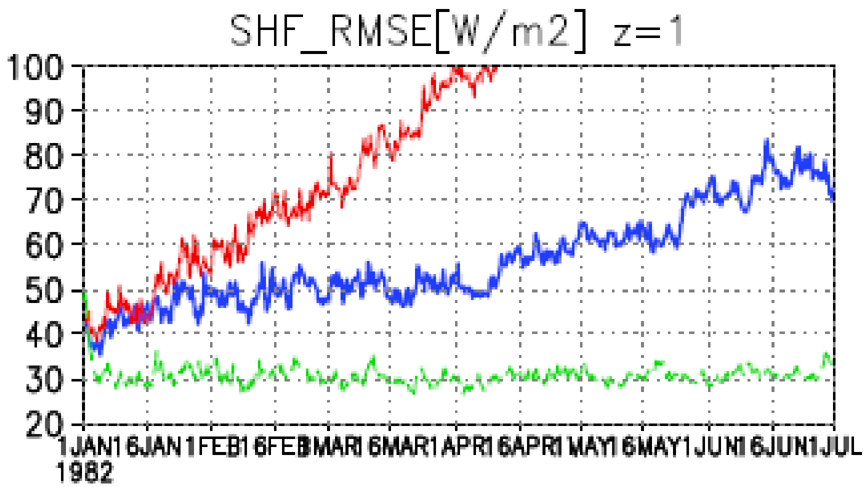


Time series of RMS errors

Analysis w/o filtering — Analysis w/ filtering — Analysis with perfect WSTR





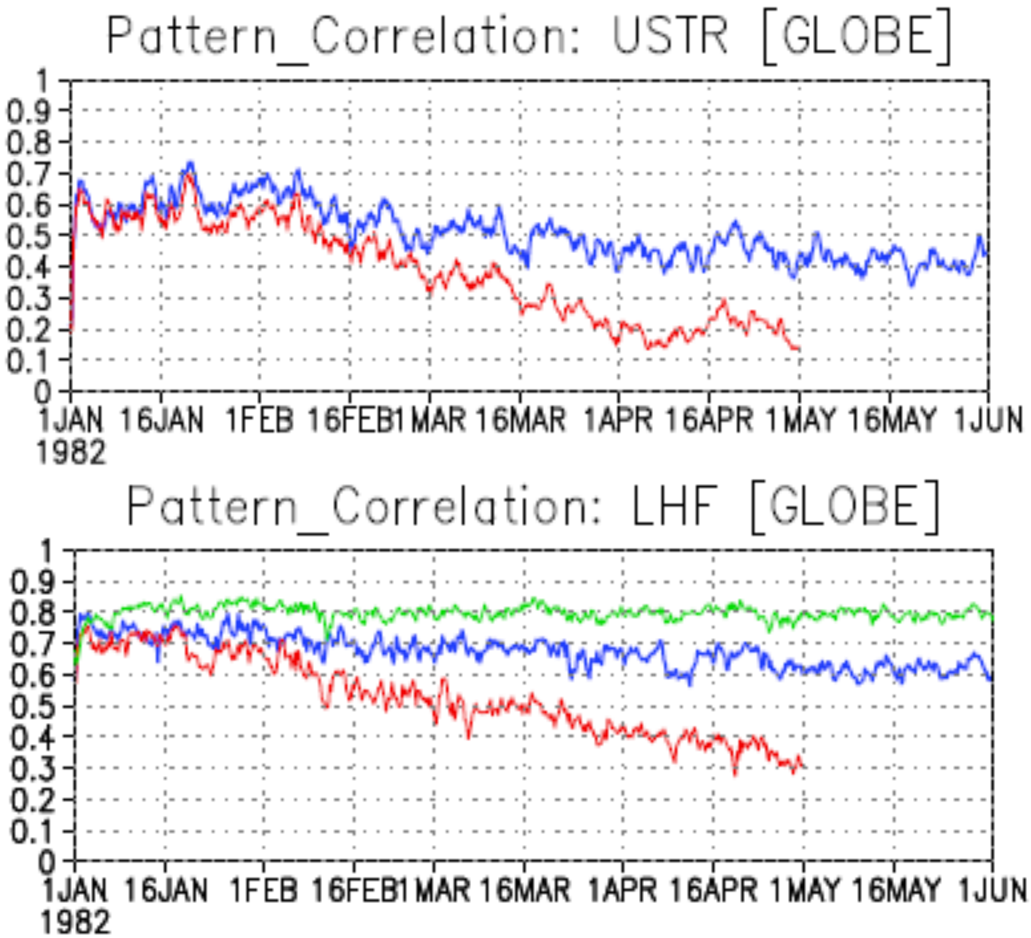


Filtering analysis increment reduces analysis error remarkably and produces quite stable results

However, there are still errors growing in time especially for the parameters (SHF, LHF, USTR, VSTR)

Time series of spatial correlation

Analysis w/o filtering — Analysis w/ filtering — Analysis with perfect WSTR



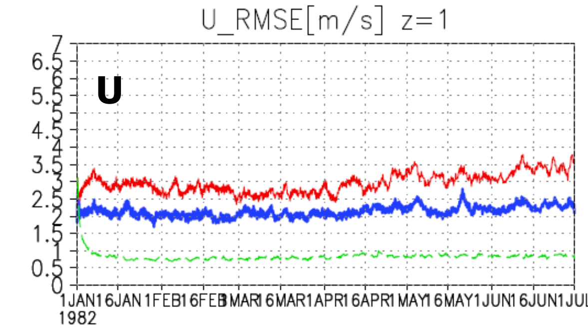
- Filtering analysis increments prevents (or delays) the estimated parameters from losing spatial correlation in time.

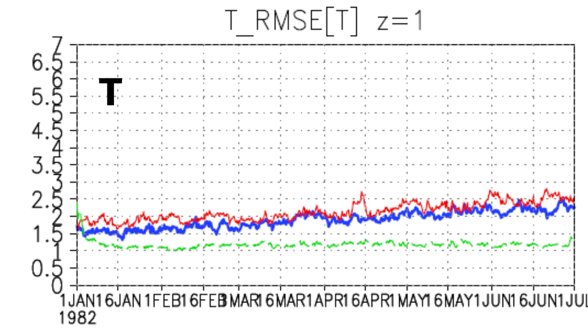
2) increasing ensemble size

- We introduced **too many unknowns** into the analysis system, and thus increasing ensemble size may help.
- Control experiments: **40** ensembles
- Experiments with **80** ensembles have been examined

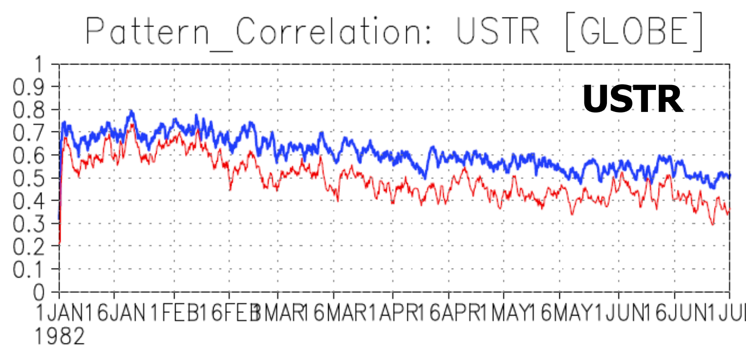
Results

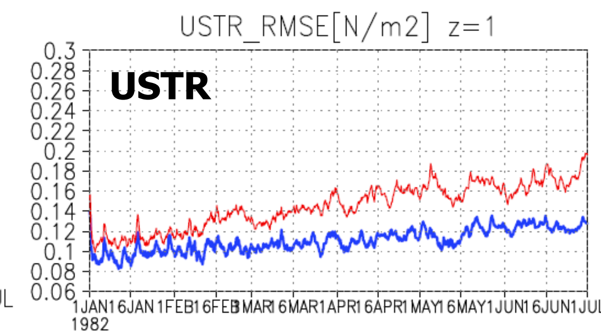
- Spatial correlation (left) and RMSE (right)
  - Blue: 80 ensembles
  - Red: 40 ensembles
  - Green: perfect WSTR with 40 ensembles

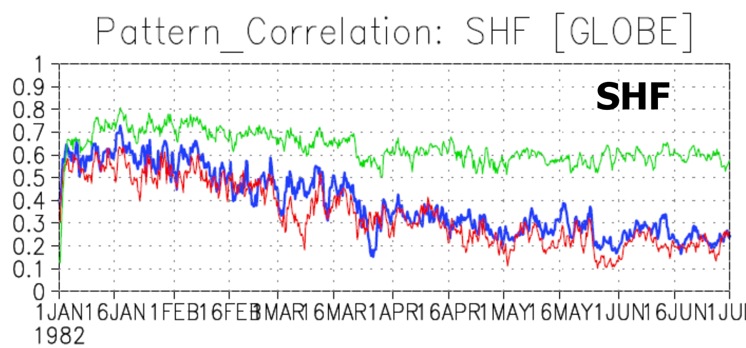


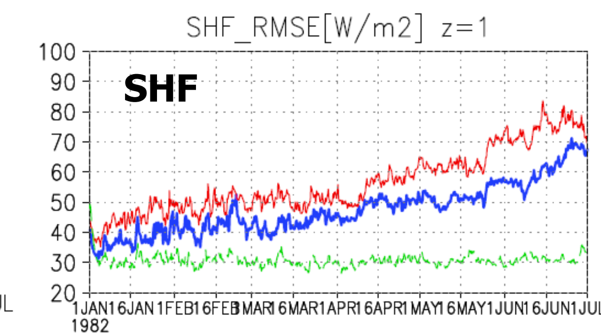


➔ Doubling ensemble size reduces error and increase spatial correlation of the estimates, but it seems not enough to produce stable estimation of parameters throughout the analysis period

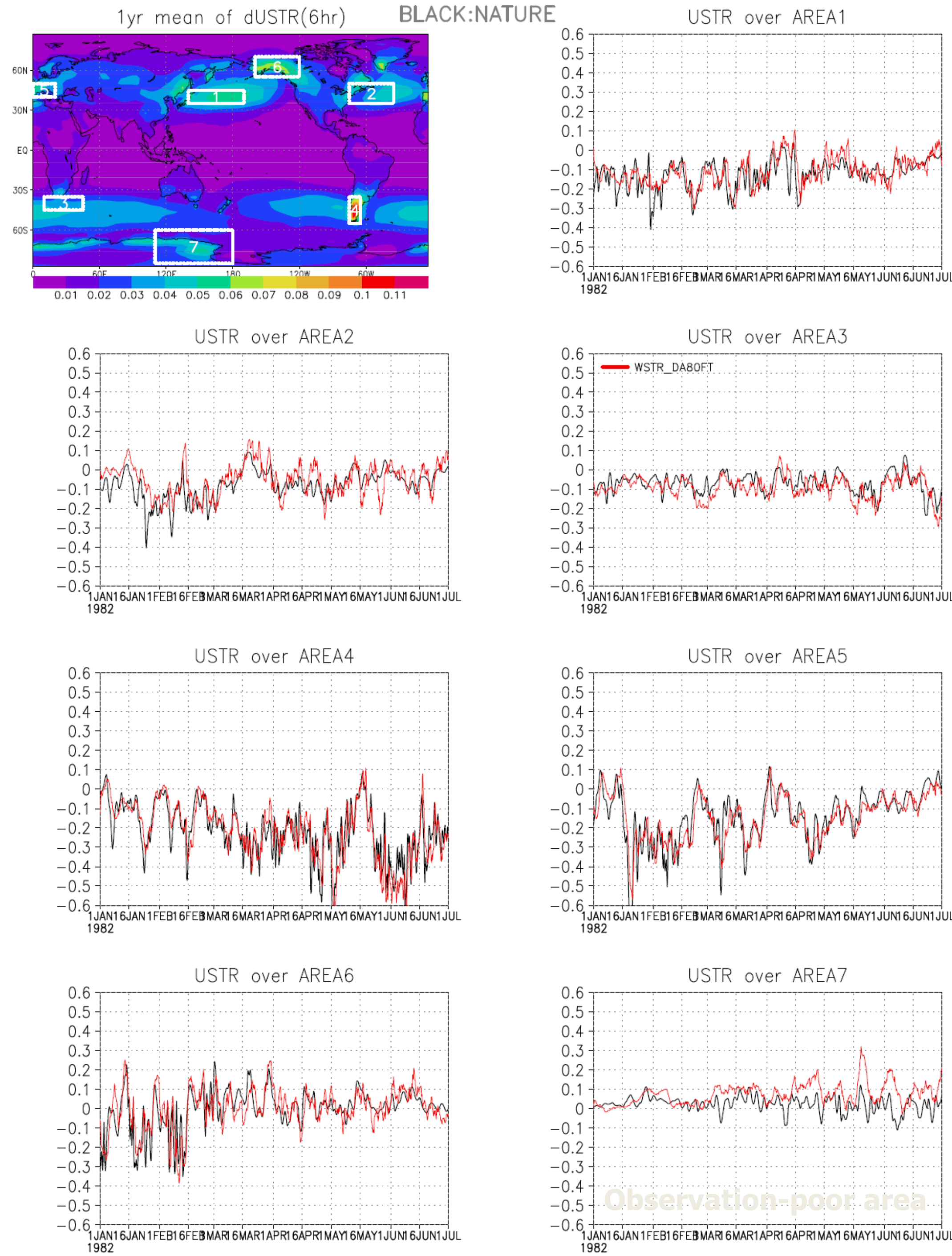




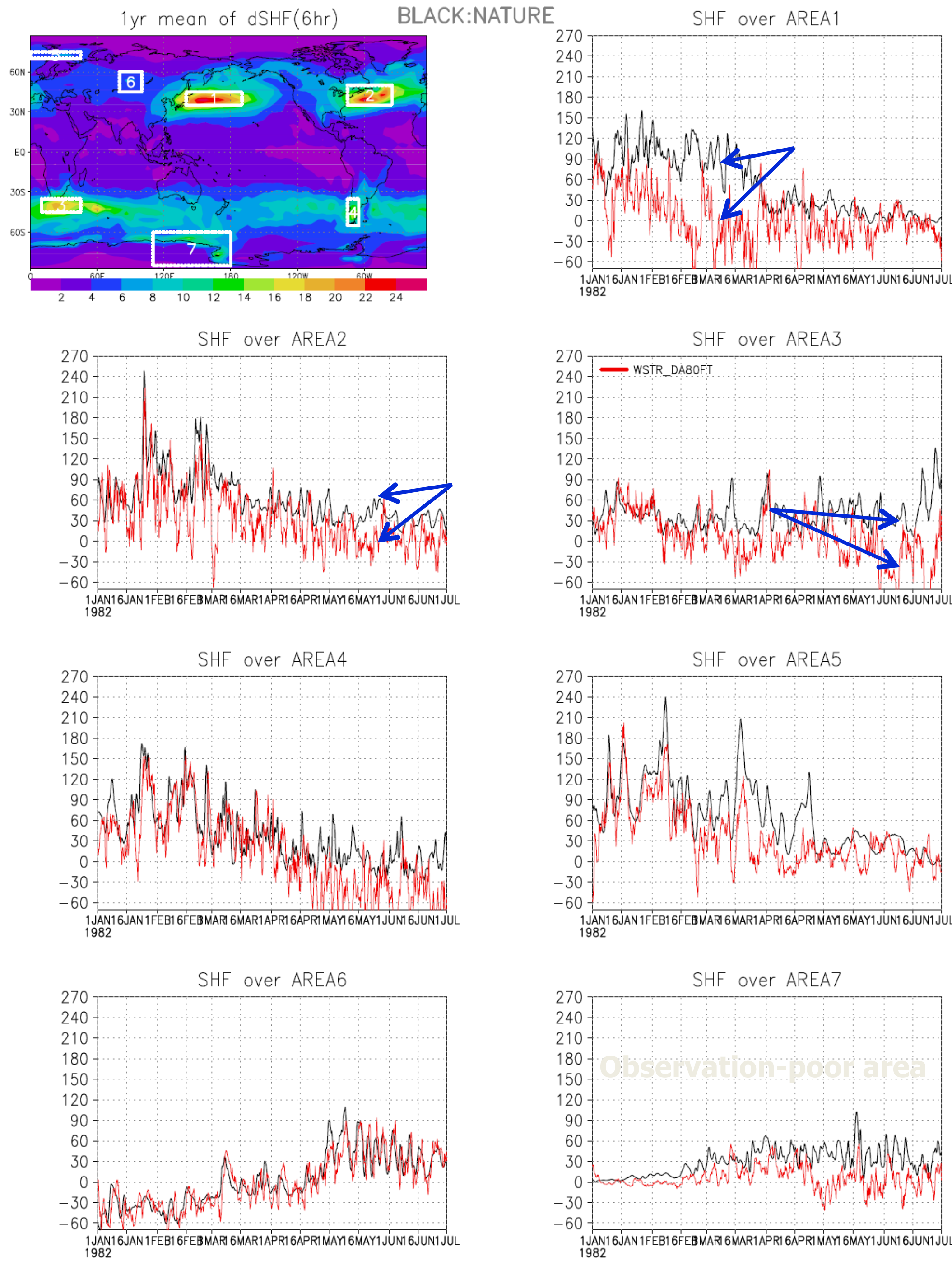




- Estimated USTR looks reasonable



- SHF tends to be underestimated, especially over the ocean
- Estimation over the land (area 4 and 6) has relatively good performance
  - Better observations over land

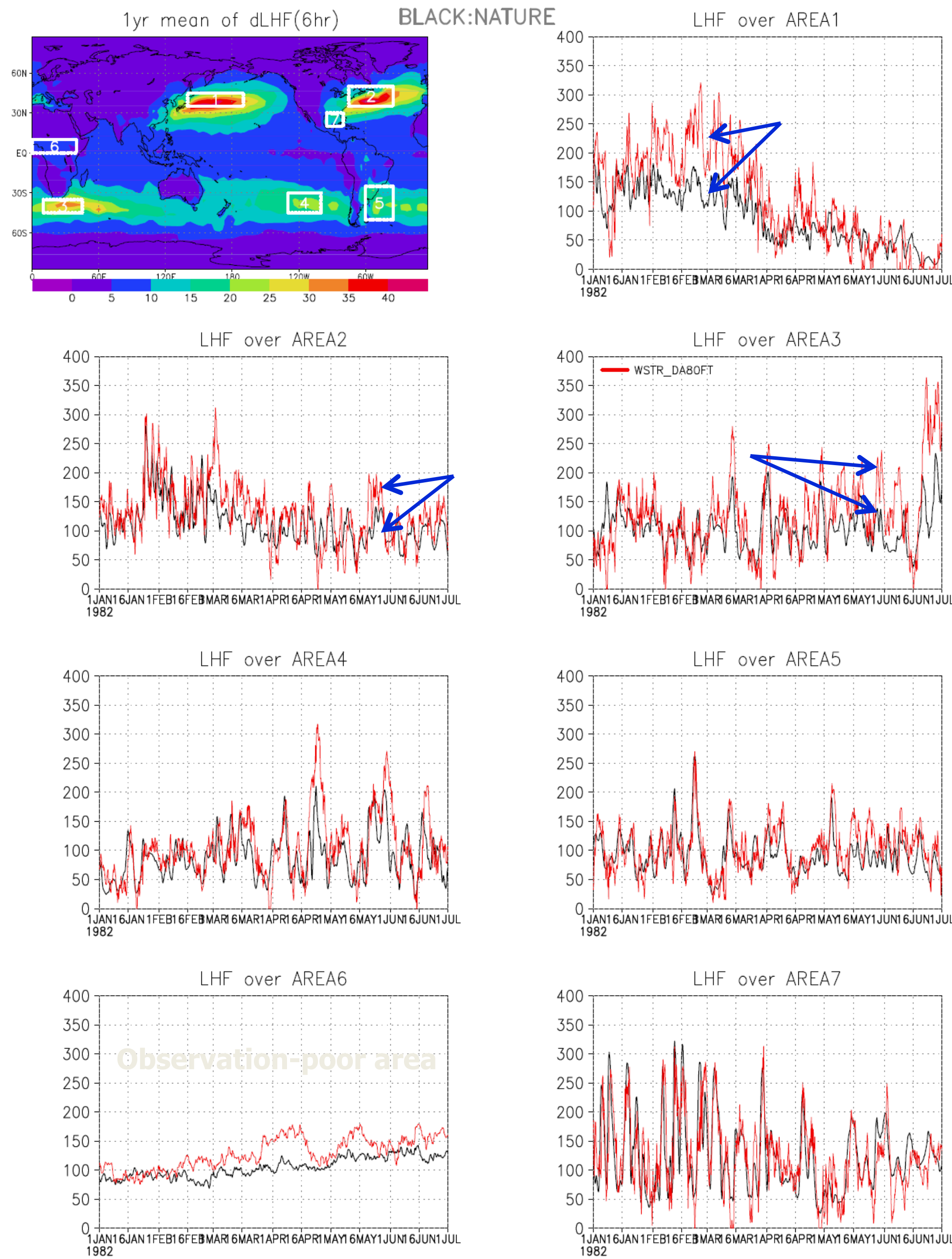


- LHF is overestimated, especially over the ocean

➔ "*improper partitioning*" *(e.g.Vinukollu et al. 2012)*

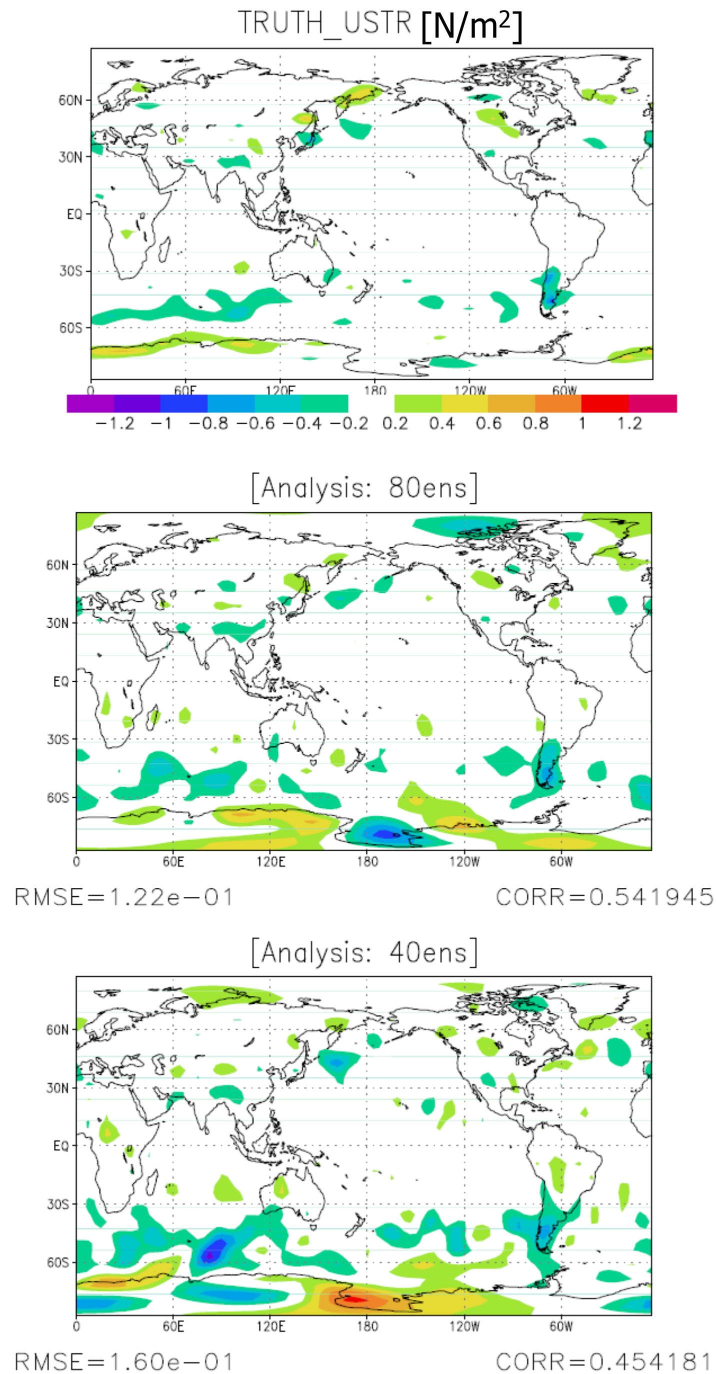
- Estimation over the land (area 5 and 7) has relatively good performance
  - Area 6 is also over the land, but there are few rawinsonde observations

➔ Results depends on the observational contents since our methods does not use any *a-priori* information



Global maps of USTR

- 00Z01JUN after a 5-month DA
- Over land, estimation of USTR agrees reasonably well with the true USTR in both experiments w/ 80 and 40 ensembles



Summary

- We have shown the feasibility of simultaneous analysis of meteorological and carbon variables within LETKF framework through OSSEs.
- The system LETKF-C has been tested in a intermediate-complexity model SPEEDY-C with excellent results.
  - Multivariate data assimilation with "localization of variables" (Kang et al. 2011)
  - Advanced data assimilation methods for CO_2 flux estimation have been explored (Kang et al. 2012)
- Implementation of the LETKF-C to NCAR CAM 3.5 model:
  - Analysis step shows very good performance in OSSE with real observation coverage
  - The same methodology has been applied to estimating surface fluxes of heat, moisture and momentum, and the results are promising!